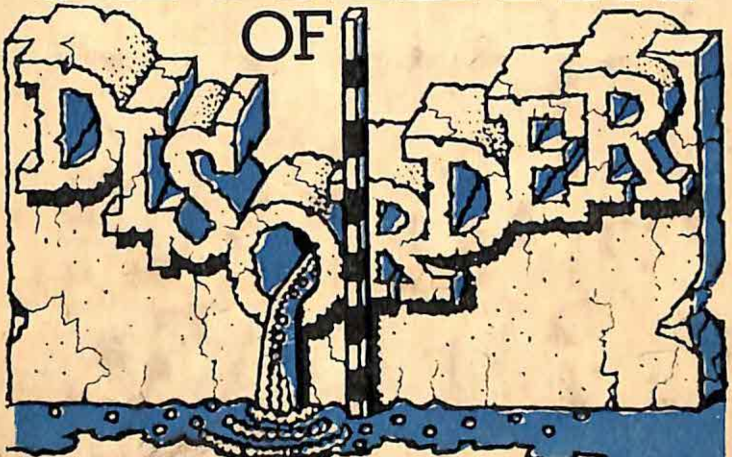
SCIENCE FOR EVERYONE

A. L. EFROS

PHYSICS AND GEOMETRY OF



PERCOLATION THEORY

MIR

Science
for Everyone

А. Л. Эфрос

Физика и геометрия беспорядка

Издательство «Наука», Москва

A. L. Efros

Physics and Geometry of Disorder

Percolation Theory

Translated from the Russian
by V. I. Kisin

S. C. B. R. T. [illegible]
Library
Calcutta

Mir
Publishers
Moscow

First published 1986
Revised from the 1982 Russian edition

S.C.E.R.T. W.B. LIBRARY
Date 7.9.98
Accn. No 9721

530
EFR

На английском языке

© Издательство «Наука». Главная редакция физико-математической литературы, 1982
© English translation, Mir Publishers, 1986

Contents

Preface 9

Part I. Site Percolation Problem . . 14

Chapter 1. Percolation Threshold 14

Two Pundits Shred a Wire Mesh (14). What Is a Random Variable? (17). Mean Value and Variance (18). Why a Large Wire Mesh? (23). Exercises (27).

Chapter 2**. Basic Rules for Calculating Probabilities. Continuous Random Variables 28

Events and Their Probabilities (28). Addition of Probabilities (30). Multiplication of Probabilities (33). Exercises (37). Percolation Threshold in a 2 × 2 Network (37). Exercise (40). Continuous Random Variables (40). Exercise (44). Percolation Threshold as a Continuous Random Variable (44). Exercise (48).

Chapter 3. Infinite Cluster 48

Permanent Magnet (48). Doped Ferromagnetics (53). Formation of an Infinite Cluster (56). Exercise (59). Site Percolation Problem Revisited (59). Clusters at a Low Concentration of Magnetic Atoms** (63). Exercises (67).

Chapter 4. Solution of the Site Percolation Problem by Monte Carlo Computer Techniques 68

Why Monte Carlo? (68). What Is the Monte Carlo Method? (70). How to Think Up a Random Number (74). The Mid-Square Method (76). Exercises (78). Linear Congruent Method (78). Exercises (79). Determination of Percolation Threshold by Monte Carlo Simulation on a Computer. Distribution of Blocked and Nonblocked Sites (81). Exercise (84). Search for Percolation Path (85). Determination of the Threshold (86). Exercise (89).

Part II. Various Problems of Percolation Theory and Their Applications 90

Chapter 5. Problems on Two-Dimensional Lattices 90

We Are Planting an Orchard (the Bond Problem) (90). Exercise (95). Inequality Relating x_b to x_s (95). Exercise (98). Covering and Containing Lattices (98). "White" Percolation and "Black" Percolation (105). Dual Lattices (110). Exercise (115). Results for Plane Lattices (116). Exercise (117). Directed Percolation (117).

Chapter 6. Three-Dimensional Lattices and Approximate Evaluation of Percolation Thresholds 120

Three-Dimensional Lattices (121). Percolation Thresholds for 3D Lattices (126). Factors Determining Percolation Threshold in the Bond Problem (127). How to Evaluate Percolation Threshold in the Site Problem (129). Exercise (134).

Chapter 7. Ferromagnetics with Long-Range Interaction. The Sphere Problem 135

Ferromagnetics with Long-Range Interaction (136). Exercise (140). The Sphere (Circle) Problem (140). The Circle (Sphere) Problem Is the Limiting Case of the Site Problem (144).

Chapter 8. Electric Conduction of Impurity Semiconductors. The Sphere Problem . . . 147

Intrinsic Semiconductors (147). Impurity Semiconductors (150). Transition to Metallic Electric Conduction at Increased Impurity Concentrations (158). The Mott Transition and Sphere Problem (161). Exercise (166).

Chapter 9. Various Generalizations of the Sphere Problem 166

Inclusive Figures of Arbitrary Shape (166). The Ellipsoid Problem (169). Other Surfaces (173). Another Experiment at the House Kitchen. The Hard-Sphere Problem (174).

Chapter 10. Percolation Level 179

"The Flood" (179). How to Construct a Random Function** (182). Analogy to the Site Problem** (185). Percolation Levels in Plane and Three-Dimensional Problems** (186). Impurity Compensation in Semiconductors (189). Motion of a Particle with Nonzero Potential Energy (190). Motion of an Electron in the Field of Impurities (192).

Part III. Critical Behavior of Various Quantities Near Percolation Threshold. Infinite Cluster Geometry 195

Chapter 11**.The Bethe Lattice . . . .

Rumors (196). Solution of the Site Problem on the Bethe Lattice (200). Discussion (204). Exercise (206).

Chapter 12. Structure of Infinite Clusters . . 206

The Shklovskii-de Gennes Model (206). Role of the System's Size (210). Electric Conduction Near Percolation Threshold (215). Exercise. (219). Function $P(x)$ Near Percolation Threshold. Role Played by Dead-Ends (219). Universality of Critical Exponents (222).

Chapter 13. Hopping Electric Conduction . . 226

Mechanism of Hopping Conduction (227). Resistor Network (229). Properties of Resistor Network (231). The Sphere' Problem Revisited (232). Calculation of Resistivity (233). Discussion of the Result (235).

Chapter 14. Final Remarks 237

Some Applications (237). What Is Percolation Theory, After All? (240).

Answers and Solutions 242

Chapter 1 (242). Chapter 2 (244). Chapter 3 (246). Chapter 4 (249). Chapter 5 (250). Chapter 6 (256). Chapter 7 (257). Chapter 8 (257). Chapter 11 (257). Chapter 12 (258).

Preface

The area of science to which this book is devoted is very young. Its basic ideas were formulated as recently as 1957 in a paper by two British scientists, S. R. Broadbent and J. M. Hammersley. During the mid-1950's Broadbent was working at the British Coal Utilization Research Association on the design of gas masks for use in coal mines. He came across an interesting problem and presented it to the mathematician Hammersley.

The principal element of a mask is filled with carbon granules through which the gas must flow. Carbon contains pores that are connected together in an intricate manner to form a sort of complicated maze. A gas can enter the pores by being adsorbed on their inner surface. It was found that if the pores are wide and well connected, the gas penetrates deep into the carbon filter. Otherwise the gas cannot get beyond the outer surface of the carbon. The motion of a gas through the maze is a new type of process, which differs from diffusion.

Broadbent and Hammersley called the phenomenon *percolation*, and the theory underlying processes of this type is referred to as *percolation theory*.

Since Broadbent and Hammersley published their pioneer paper 28 years ago, it was discovered that percolation theory can be used to interpret an exceptionally wide variety of physical and chemical phenomena. The electrical properties of disordered systems, such as amorphous semiconductors, crystalline semiconductors with impurities, and materials formed as mixtures of a dielectric and a metal, are probably the best understood application of percolation theory.

The phenomena best described by percolation theory are *critical phenomena*. They are characterized by a *critical point* at which some of the properties of the system undergo abrupt changes. Critical phenomena include second-order phase transitions (e.g. the transition of a metal from its normal to its superconducting phase when its temperature is lowered). The physics of all critical phenomena is very unusual, but there are some common features, the most important of which is that in the neighborhood of the critical point, the system appears to break into blocks which differ in their properties, with the size of the individual blocks growing until the system approaches the critical point. The blocks are quite randomly shaped. In some phenomena the whole configuration changes chaotically because of thermal motion, while in other phenomena the configuration may be frozen in time but changes from specimen to specimen. The blocks are in complete disorder, so that no regularity is discernable from an instantaneous photograph. However, this geometry, which can be called the *geometry of disorder*, has quite definite properties "on the average".

Actually, geometry is inseparable from physical properties. For instance, the physical properties of a crystal are determined by the geometry of its lattice. Likewise, the "geometry of disorder" determines a number of properties of a system in the vicinity of a critical point. The most interesting feature is that owing to the large size of the blocks the geometry is virtually independent of the atomic structure of the material and thus possesses properties common to a number of quite dissimilar systems; hence, the universality of the physical properties that we find in the neighborhood of critical points.

This type of relation between physics and geometry can be traced in percolation theory, and in fact, making this relationship clear is the main objective of this book. Percolation theory is formulated in terms of simple geometric images, such as wire nets, spheres or crystal lattices. The theory does not operate with the concept of temperature, and this makes it possible to clarify the idea of critical phenomena for readers not familiar with statistical physics.

Percolation theory, as a theory of critical phenomena, is not yet a mathematically rigorous science. A large number of important propositions have not yet been proved, and certain questions have not been answered. However, if a rigorous proof exists but appears to be complicated, I decided to replace it here by arguments that will explain the result rather than prove it. However, an effort was made to separate those propositions that have not yet been proved from those that have.

The book offers a detailed presentation of the theory of percolation and its various applications. However, a definition of what is understood by percolation theory and what processes it describes is deferred until the last page of the book. The definition must encompass so many complex concepts that to give it at the beginning would be meaningless. Nearly every chapter concentrates on one specific problem whose analysis leads to a problem in percolation theory. It is assumed that, having read several chapters, the readers will become aware of the common ground the various percolation problems have, and also become aware of the relevance of the title of the book.

As a rule, the problems I chose represent important applications of percolation theory. However, some (such as laying out an orchard in Chapter 5 and the propagation of rumors in Chapter 11) are illustrative or even slightly humorous.

The fundamentals of elementary probability theory required for understanding the material are also covered. Chapter 1 gives a general notion of probability and random variables. Chapter 2 introduces the rules for the addition and multiplication of probabilities and defines the distribution function. An "easy reading" version of the book is obtained by skipping Chapter 2 and the other chapters and sections marked with two asterisks, although the reader will not then be able to follow the derivation of certain of the quantitative results or a few of the exercises. Nevertheless, this should not hamper the understanding (even if slightly less comprehensive) of the other chapters.

I consider the exercises in the text important. As a rule, the exercises are quite simple, and I recommend that the reader do them without looking at the answers section beforehand unless this is specifically advised.

B. I. Shklovskii played an important role in the creation of this book because together we discussed its structure and title, and then he read the manuscript. I am extremely grateful to him for his contribution.

I am also grateful to my colleagues L. G. Aslamazov, N. B. Vasilyev, Yu. F. Berkovskaya, and M. E. Raikh because they read the manuscript and suggested a number of useful changes.

I am especially grateful to my wife N. I. Efros for carrying the heavy burden of preparing the manuscript for publication.

A. Efros

Part I

Site Percolation Problem

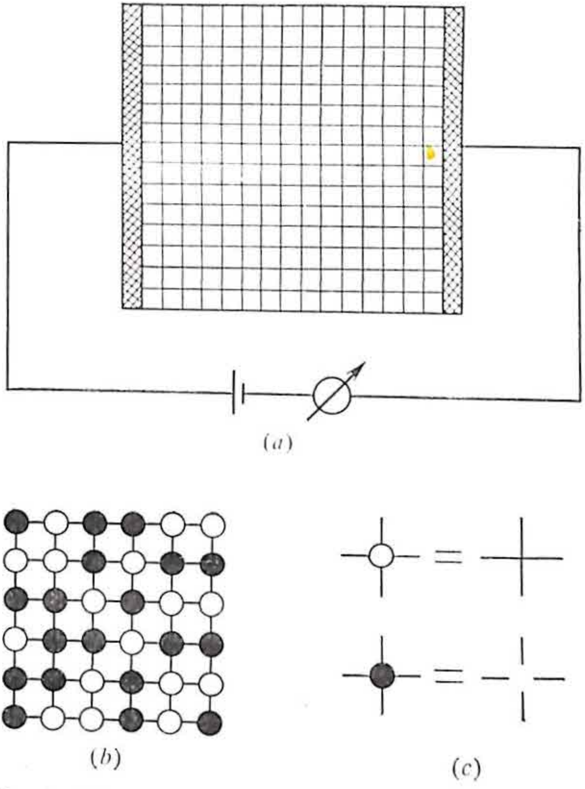
Chapter 1

Percolation Threshold

Two Pundits Shred a Wire Mesh

It is not frequent nowadays for a scientific journal to publish a report on experiments done with, for example, a piece of ordinary wire mesh bought in the nearest hardware store. A paper written by two American physicists Watson and Leath, published in the *Physical Review* in 1974, was definitely not the first one in the realm of percolation theory; nevertheless, it is with this study that we start our story.

The piece of wire mesh used by Watson and Leath was a square containing $137 \times 137 = 18\,769$ nodes, or sites, with the neighbor sites separated by a spacing of $1/4$ inch $= 6.35$ mm. The scientists soldered copper electrodes to two opposite sides of the square and connected the mesh to an electric resistance-measuring circuit (Fig. 1*a*). Then they started to block some sites in the mesh and monitored the electric resistance as a function of the fraction of blocked sites. As



(a)

(b)

(c)

Fig. 1. Schematic of the experiment of Watson and Leath: (a) an initial wire mesh (the number of sites in the figure is greatly reduced); (b) a piece of wire mesh with blocked sites (blocked sites are shown by black circles, and non-blocked sites by open (white) circles); (c) a black site signifies that the four wires coming to the site are not in contact, and a white site signifies intact contacts (no electric current flows through black sites in any direction, while it flows through white sites in any direction).

shown in Fig. 1*b*, *c*, a site was blocked by simply cutting all four wires joining at this site.

The choice of a new site to be blocked was made randomly among those still in the mesh. In principle, this could be done by writing the coordinates of each site on separate slips of paper, putting all these slips into a hat, stirring well the contents, and then extracting them one by one. However, such a procedure, as well as any other mechanical method of sequence allotment, is highly impractical if the number of sites is large; consequently, the scientists used a random sequence of site coordinates generated in a computer. Later we shall describe how a computer can be "made to generate" random numbers, but for the time being we can think that a computer was replaced, with no loss to clarity, by a hat-and-paper-slips combination.

Obviously, the electric conductance of the mesh decreased with increasing number of blocked sites. (Electric conductance is defined as a quantity reciprocal of resistance. Resistance is measured in ohms, and electric conductance in inverse ohms (ohm^{-1}).) Furthermore, if x denotes the ratio of nonblocked sites to the total number of sites (137^2), the electric conductance vanished at a certain value of x that hereafter we call the *threshold (critical) value*, or *percolation threshold*, and denote by x_c. This vanishing occurred when the last path connecting the left- and right-hand electrodes had been cut. One of the objectives of the experiment was to determine x_c. It was found that $x_c = 0.59$.

Probably, the first question that must be answered is whether the variable x_c is random, irreproducible from one experiment to another, or x_c is quite definite. Let us assume that the experiment is repeated with another square piece of wire mesh and with a different random sequence of blocked sites. It is common sense to expect that the configuration of blocked and intact sites being quite different at each stage of the second experiment in comparison with the first, the cutting of the last path connecting the electrodes must also occur at a different value of x, so that the second experiment must give a new x_c. This is certainly correct.

The threshold value x_c in the experiment we are now discussing is a *random quantity*, or *variable*. Quantities of this sort will be encountered throughout the book, and so it is useful to answer from the very beginning the following question:

What Is a Random Variable?

In mathematics a variable is said to be random if the values it takes and the frequency at which it takes them are known, but what particular value the variable will assume in each particular case is unknown (and cannot be known in the framework of the given mathematical problem).

Here is a classical example of a random variable: a cube (a die) with numbered faces is being thrown onto a table. The numeral read off the topmost face is a random variable. Such a variable is said to be *discrete* because it assumes only certain discrete values (in the chosen example these are six numerals: 1, 2, 3, 4, 5, 6). It would be impossible to predict the specific numeral obtained in each given trial (i.e. in each throw), but the *probability* of getting a specific number (e.g. 4) is predictable. Assume that the number of trials was Q, and face 4 appeared in Q_4 cases. The ratio Q_4/Q is called the *relative frequency of recording a given value of the random variable* (here, the numeral 4). If the total number of trials is not very large, this ratio varies: in another series of Q trials the ratio Q_4/Q may be substantially different. However, as the number of trials Q increases, these fluctuations progressively diminish. The relative frequency of recording a given value of the random variable approaches the limit which is called the *probability* of this value.

Let us denote the probability of obtaining face 4 by P (4). If the cube is "honest" (not "loaded"), that is, if all its faces are equivalent, the quantity P (4) is easily predicted. On the average, any one of the six faces of the cube will turn up topmost the same number of times, so that $Q_4/Q = Q_3/Q = \ldots = 1/6$ provided Q is large. Therefore, the probabilities of recording the six numerals are identical and equal to 1/6.

We thus find that after a large number of throws, chance melts into the background and clears the stage for a regularity, namely, the symmetry of cube's faces.

Mean Value and Variance

Let us return to the experiment with the wire mesh. We concluded that since the experiment operated with a random sequence of blocked sites, the critical concentration x_c at which the current between the left- and right-hand electrodes is interrupted is also a random variable, and it would be impossible to predict the value of x_c in each specific experiment.

The theoretical approach to the situation could be a study of the "mean" properties of the quantity x_c, that is, of the properties apparent after a sufficiently large number of experiments run under identical conditions. These conditions are, first, the total number of sites in the wire mesh, $\mathcal{N}$ ($\mathcal{N} = 137^2$ in the experiment described above), and second, the properties of the generator of random numbers that prescribes a random sequence of blocked sites. The requirement that the properties of the generator should not change from one experiment to another does not mean at all that the sequences of blocked sites must be identical. (In this case all values of x_c would be identical!) All we need is to use the same method of generating the random sequence of sites to be blocked in all experiments (e.g. a hat with slips of paper).

Having conducted Q experiments with a wire mesh containing $\mathcal{N}$ sites, we obtain Q values x_i, where the subscript i is the number of the experimental run. For example, x_{15} denotes the x_c obtained in the fifteenth series of measurements. The most important of the mean values is the arithmetic mean, or the average, $\bar{x}_Q$, obtained by adding up all x_i and dividing the sum by the number Q of experiments:

$$\bar{x}_Q = \frac{x_1 + x_2 + \ldots + x_Q}{Q} \qquad (1)$$

The bar over x and the subscript Q denote averaging over the results of Q experiments. The quantity $\overline{x}_Q$ is still a random variable. If we conduct another series of Q experiments under the same conditions and calculate the resultant $\overline{x}_Q$, this $\overline{x}_Q$ will be somewhat different. However, the larger the number of experiments, Q, within the series, the smaller the differences between the averages representing different series. The point is that *random fluctuations of x_i cancel each other in the long run, so that as Q increases, the arithmetic mean $\overline{x}_Q$ tends to a quite definite value independent of Q but dependent on the conditions under which the experiments were run*. This limiting value is called the *mean value of a random variable*. (In probability theory this limiting value is also called the *expectation* of a random variable, but this term will not be used in this book.)

The mean value of the percolation threshold of a wire mesh made up of $\mathscr{N}$ sites will be denoted by $x_c(\mathscr{N})$. The quantity $x_c(\mathscr{N})$ is not a random variable but a *certain quantity*. Its dependence on $\mathscr{N}$ is a regular property worth thinking about.

An important characteristic of a random variable x_c is also found in the deviations δ_i of the values x_i from the mean value:

$$\delta_i = x_i - x_c(\mathscr{N}) \tag{2}$$

The deviations δ_i vary from one experiment to another, and we want to select a quantity characterizing the properties of δ_i "on the average". We cannot select the arithmetic mean for this quantity because it tends to zero when $Q \to \infty$. Indeed,

$$\frac{\delta_1+\delta_2+\ldots+\delta_Q}{Q}=\frac{x_1+x_2+\ldots+x_Q}{Q}-x_c(\mathcal{N})$$

But the greater Q is, the less the difference is between the first term on the right-hand side of this equality and the second term; this proves the proposition made above. This result is obtained because the individual values of x_i necessarily fall on both sides of the mean value, so that on the average the deviations cancel out.

We could also take the arithmetic mean of the nonnegative quantity $|\delta_i|$; however, the standard procedure is to calculate the *variance* $\delta^2(\mathcal{N})$ which is the arithmetic mean of the *squares* of deviations (for $Q \to \infty$) which, of course, are also nonnegative quantities:

$$\delta^2(\mathcal{N})=\frac{\delta_1^2+\delta_2^2+\ldots+\delta_Q^2}{Q} \tag{3}$$

The quantity $\delta(\mathcal{N}) = [\delta^2(\mathcal{N})]^{1/2}$ is called the *root-mean-square deviation* of a random variable, or rms *deviation*. It is the rms deviation $\delta(\mathcal{N})$ that characterizes the typical deviation of the values x_i from their mean value $x_c(\mathcal{N})$. Obviously, the quantity $\delta(\mathcal{N})$ is also a function of the total number of sites, $\mathcal{N}$, in the wire mesh.

Strictly speaking, x_c is a discrete quantity because it is obtained by dividing the number of nonblocked sites by the total number of sites, $\mathcal{N}$, and therefore, it assumes only those values that convert to integers when multiplied by $\mathcal{N}$. Let us denote all possible distinct values of the random variable x_c by x_k.

7.9.98
9721

The mean value $x_c(\mathcal{N})$ can be expressed in terms of the probability $P(x_k)$ of the random variable x_c assuming the value x_k. Remember that expression (1) contains the sum of x_i obtained in Q experiments. Each value of x_i can appear many times. Formula (1) can be written in the form

$$\overline{x}_Q = \frac{x_1 Q_1 + x_2 Q_2 + \ldots}{Q} \tag{4}$$

where the summation is carried out over all *distinct* values x_k that x_c can assume (no one value of x_k is encountered twice in this sum!). The factor Q_k is equal to the number of times the value x_k turned up in a series of Q experiments.

The quantity Q_k/Q is a relative frequency of obtaining x_k as a result. When Q is very high, this ratio becomes equal to the probability $P(x_k)$. By definition, for large values of Q the left-hand side of expression (4) turns into $x_c(\mathcal{N})$. Consequently,

$$x_c(\mathcal{N}) = x_1 P(x_1) + x_2 P(x_2) + \ldots \tag{5}$$

i.e. the mean value equals the sum of all the values that a random variable can assume, multiplied by their probabilities. Likewise,

$$\delta^2(\mathcal{N}) = (x_1 - x_c(\mathcal{N}))^2 P(x_1) + (x_2 - x_c(\mathcal{N}))^2 P(x_2) + \ldots \tag{6}$$

The summation in formulas (5) and (6) is carried over all possible values that the random variable x_c can assume, with each value encountered only once in the sum.

By definition, $Q_1 + Q_2 + \ldots = Q$, so that the definition of $P(x_k)$ yields

$$P(x_1) + P(x_2) + \ldots = 1 \qquad (7)$$

The sum of the probabilities of all the values that a random variable can assume equals unity.

Why a Large Wire Mesh?

It was very simple to calculate, in the problem of "honest" cube, the probability for the random variable to assume a specific value. The properties of the random variable x_c are incomparably more complex.

It will be shown at the end of the next chapter how to solve the problem for a square network of four nodes (sites) (2×2 sites, $\mathscr{N} = 4$). The result of the solution: the random variable x_c can assume only two values, namely, 1/2 and 1/4. The first of these is assumed with probability $P(1/2) = 2/3$, and the second with probability $P(1/4) = 1/3$. According to formulas (5) and (6) (where the sums consist of only two terms), $x_c(4) = 5/12$ and $\delta(4) = \sqrt{2}/12$.

With sufficient patience, we can also solve the problem with 3×3 sites ($\mathscr{N} = 9$). Experience shows that the required efforts increase enormously if the side of the square is augmented by only one site. At the same time, the networks whose properties are of especially high interest contain a very large number of sites (e.g. 10^{15}). Such networks can serve as models of films consisting of atoms. Indeed, the spacings between the atoms in condensed materials (liquids and

S.C.E.R.T. W.B LIBRARY
Date

crystals) are as a rule about $3 \cdot 10^{-8}$ cm. Consequently, a film covering 1 cm^2 with a thickness of a single layer of atoms consists of roughly 10^{15} atoms.

The problem of finding the probability for the percolation threshold of a network made up of a very large number of sites, $\mathscr{N}$, to assume a specific value is the central problem of percolation theory. It will be discussed, in one form or another, throughout the book. Here we want to single out, giving practically no proof, the most important property of this problem that offers the key to understanding the problem as a whole:

> The root-mean-square deviation $\delta(\mathscr{N})$ diminishes with increasing number $\mathscr{N}$ of sites in power-law fashion, tending to zero as $\mathscr{N} \to \infty$.

This property is expressed by the formula

$$\delta(\mathscr{N}) = \frac{C}{\mathscr{N}^{1/2\nu}} \tag{8}$$

where $C \approx 0.54$ and $\nu \approx 1.3$. (The quantity ν is called the *exponent of correlation radius*. It is discussed in detail in Part III of the book.)

Watson and Leath could not arrive at formula (8) as a result of their experiment. To obtain (8), it was necessary to use wire meshes with different values of $\mathscr{N}$ and to run numerous experiments at the same $\mathscr{N}$. Besides, expression (8) is a result of the theoretical studies to be discussed in Part III.

As follows from formula (8), the results of experiments with different random sequences of blocked sites differ from one another the less, the greater the number of sites in the wire mesh.

Why should this be so? The point is that a sufficiently large wire mesh manifests all or nearly all the possible configurations of intact and blocked sites. These configurations appear to change places in different experimental runs. Consequently, the role of chance is the smaller, the greater $\mathscr{N}$ is. An infinite net contains an infinitely large number of large subnets, so that randomness ceases to play any role at all, and *the value of x_c is not a random variable but a certain quantity equal to*

$$x_c = \lim_{\mathscr{N} \to \infty} x_c(\mathscr{N})$$

It is this limiting value that is in fact defined as percolation threshold, and it was for finding this threshold that Watson and Leath have conducted their experiment. Otherwise why should they have taken a wire mesh with nearly 19 000 sites? They could have taken a 2 × 2 wire mesh!

Now let us formulate the most important result of this chapter:

> The concept of a sharply defined percolation threshold independent of the choice of a random sequence of blocked sites used in the experiment holds in an infinitely large system. No sharply defined threshold exists in a finite system, but there is the so-called critical region with width of the order of $\delta(\mathscr{N})$ into which the values of x_c, which are recorded in the majority of experimental runs with different random sequences, fall. As the size of the system increases, this region contracts to a point.

It must be borne in mind, however, that the dependence on the size of the system is important only as long as we attempt to model a phenomenon artificially (e.g. by using a wire mesh). As a rule, percolation theory is applied to systems in which individual elements are extremely small (e.g. atoms, for as we have already mentioned (see Chapter 3) 1 cm^2 of a monatomic layer contains $\mathcal{N} = 10^{15}$ elements, and 1 cm^3 contains $\mathcal{N} = 10^{22}$!), and so to a very good accuracy, the systems can be considered infinite, and the uncertainty in percolation threshold due to a system's size can be ignored.

The problem that Watson and Leath were solving is referred to as the *site percolation problem*, or simply *site problem* (because the random elements are represented by sites). Quite a few problems in science reduce to site percolation, and one of them (a doped ferromagnetic) is discussed in Chapter 3.

The exact value of percolation threshold for this problem has not been found yet. The quantity $x_c(\mathcal{N})$ is found approximately for large values of $\mathcal{N}$ in computer simulations or in so-called analogue experiments similar to that of Watson and Leath. (The techniques involved may vary substantially.)

The degree to which the result obtained deviates from the sought limiting value can be evaluated from the change in $x_c(\mathcal{N})$ with changing $\mathcal{N}$. A comparison of results obtained by different methods makes it possible to conclude that the number 0.59 is correct to within two decimal places (although it is far from obvious beforehand that the number of sites $\mathcal{N} = 137^2$ is sufficient).

Obviously, there can be no end to improving the accuracy of x_c in the subsequent decimal places (see Exercise 4).

Exercises

1. Define a discrete random variable a as the numeral on the topmost face of a cube after a throw. Find the mean value of a.

2. Define percolation threshold as a value of x at which percolation sets in not from left to right, but from top to bottom. Will this change the results of individual experimental runs, $x_c(\mathcal{N})$, x_c? Assume the network to be square.

3. Answer the same question but take into account that percolation threshold is defined as the minimum value of x at which percolation exists both from left to right and from top to bottom.

4. Answer the same question but in the case when percolation threshold is defined as the maximum value of x at which there is no percolation both from left to right and from top to bottom.

5. Make use of formula (8) and calculate the root-mean-square deviation in the conditions of the Watson-Leath experiment ($\mathcal{N} = 137^2$). What accuracy can be expected if only one experimental run was carried out?

Note. In principle, the result of a single experiment may differ very much from the mean value $x_c(\mathcal{N})$. However, using the distribution function given below for percolation thresholds (see formula (6) in Chapter 2), it can be proved that the probability for the result of a randomly se-

It must be borne in mind, however, that the dependence on the size of the system is important only as long as we attempt to model a phenomenon artificially (e.g. by using a wire mesh). As a rule, percolation theory is applied to systems in which individual elements are extremely small (e.g. atoms, for as we have already mentioned (see Chapter 3) 1 cm^2 of a monatomic layer contains $\mathcal{N} = 10^{15}$ elements, and 1 cm^3 contains $\mathcal{N} = 10^{22}$!), and so to a very good accuracy, the systems can be considered infinite, and the uncertainty in percolation threshold due to a system's size can be ignored.

The problem that Watson and Leath were solving is referred to as the *site percolation problem*, or simply *site problem* (because the random elements are represented by sites). Quite a few problems in science reduce to site percolation, and one of them (a doped ferromagnetic) is discussed in Chapter 3.

The exact value of percolation threshold for this problem has not been found yet. The quantity $x_c(\mathcal{N})$ is found approximately for large values of $\mathcal{N}$ in computer simulations or in so-called analogue experiments similar to that of Watson and Leath. (The techniques involved may vary substantially.)

The degree to which the result obtained deviates from the sought limiting value can be evaluated from the change in $x_c(\mathcal{N})$ with changing $\mathcal{N}$. A comparison of results obtained by different methods makes it possible to conclude that the number 0.59 is correct to within two decimal places (although it is far from obvious beforehand that the number of sites $\mathcal{N} = 137^2$ is sufficient).

Obviously, there can be no end to improving the accuracy of x_c in the subsequent decimal places (see Exercise 4).

Exercises

1. Define a discrete random variable a as the numeral on the topmost face of a cube after a throw. Find the mean value of a.

2. Define percolation threshold as a value of x at which percolation sets in not from left to right, but from top to bottom. Will this change the results of individual experimental runs, $x_c(\mathscr{N})$, x_c? Assume the network to be square.

3. Answer the same question but take into account that percolation threshold is defined as the minimum value of x at which percolation exists both from left to right and from top to bottom.

4. Answer the same question but in the case when percolation threshold is defined as the maximum value of x at which there is no percolation both from left to right and from top to bottom.

5. Make use of formula (8) and calculate the root-mean-square deviation in the conditions of the Watson-Leath experiment ($\mathscr{N} = 137^2$). What accuracy can be expected if only one experimental run was carried out?

Note. In principle, the result of a single experiment may differ very much from the mean value $x_c(\mathscr{N})$. However, using the distribution function given below for percolation thresholds (see formula (6) in Chapter 2), it can be proved that the probability for the result of a randomly selected experiment to fall in the interval from $x_c(\mathscr{N}) - \delta$ to $x_c(\mathscr{N}) + \delta$ is roughly 0.7. The greater $\mathscr{N}$ is, the smaller the "typical deviation" from the mean value is.

Chapter 2**

Basic Rules for Calculating Probabilities. Continuous Random Variables

This book is devoted to the laws governing disorder, and widely uses the concepts of probability and random quantity, or variable. These concepts have been partly introduced in the preceding chapter, so that the reader who is not inclined to delve into the mathematical side of the problem is welcome to skip Chapter 2, as well as all the subsequent chapters and sections marked with double asterisks. The reader, who wants to follow the solution of a number of beautiful mathematical problems given in the book and to form a more profound picture of percolation theory, must know the rules for the addition and multiplication of probabilities presented in this chapter.

Events and Their Probabilities

The concept of probability is used not only when we deal with numerical values taken up by a random variable. Any experiment with a random outcome can be discussed in these terms. The distinct results of experiments are referred to as *events*. *The relative frequency of an event is defined as the ratio of the number of trials that led to this event, to the total number of trials. The probability of an event is defined as the limit which the relative frequency of the event approaches when the number of trials tends to infinity.*

Example. Red, green, and blue balls in identical numbers are put in a box. The balls are then stirred, and one of them is extracted in a grab-bag fashion. What is the probability of the event consisting in extracting a red ball? In contrast to the experiment with dies, here events differ not in quantity but in quality (the color of balls). However, the argument runs along the same line. The number of balls being identical in each color, a red ball will be chosen in 1/3 of all trials. Consequently, the sought probability equals 1/3. The probability of drawing a blue ball or a green ball is also 1/3.

By definition, probability is a quantity that varies from zero to unity. An event with zero probability is, for instance, the extraction of a blue ball from a box with only *red* balls. The probability of drawing a red ball from the same box is unity. An event with probability equal to unity is said to be a *certain event*, not a random event.

The concept of probability plays an enormous role in clarifying the regularities in the world of random processes. A law is quite often virtually buried under chance. Imagine that you try to find a regularity in birthrates of boys and girls from the record of one maternity. You see a random sequence of the type BGBBBBGBGG....

Sometimes it seems that boys appear much more often, and sometimes your conclusion has to be reversed. Your friend assures you that "only girls are born nowadays". His impression may stem from the fact that girls were born in three families that he visits.

However, a regular pattern does exist. The ratio of the probability for a boy to be born to that for a girl to be born is 51.5 to 48.5. The regularity is observed to hold quite well in such large countries as the USSR or the USA, even if the data cover only one year.

In contrast to the problem with balls of different colors, the problem of male/female ratio in childbirth is very difficult to solve theoretically. However, the statistical data involved do reveal certain thoroughly analyzed properties of human physiology.

Addition of Probabilities

Events are said to be *incompatible* if they cannot be observed in the same trial. For instance, the event consisting in drawing a red ball is incompatible with the event consisting in drawing a blue ball because, by the conditions of the problem, only one ball can be drawn in any trial: either red, or blue, or green. The events consisting in getting the numerals 5 and 2 in one throw of our cube are incompatible.

We shall prove two important properties of probability.

1. **Addition rule.** *The probability for one (no matter which) event out of several incompatible events equals the sum of the probabilities of these*

events. Suppose that we want to find the probability for a cube to show either 3 or 4. The number of trials in which these two outcomes were obtained is equal to the number of trials that gave 3 and the number of trials that gave 4. By definition, the sought probability will be found if this sum is divided by the total number Q of trials, and the limit is found for $Q \to \infty$. The limit of each term of the sum divided by Q equals the probability of obtaining one of the numerals of interest, the sought probability indeed equals the sum of the probabilities of each of them. Therefore, the probability of obtaining either 3 or 4 is $1/6 + 1/6 = 1/3$. The probability of obtaining either 1 or 2 or 3 or 4 is $1/6 + 1/6 + 1/6 + 1/6 = 2/3$. And the probability of getting either 1 or 2 or 3 or 4 or 5 or 6 is $1/6 + 1/6 + 1/6 + 1/6 + 1/6 + 1/6 = 1$. This result is a particular case of the second property.

2. *Let us refer to the set of incompatible events that covers all possible outcomes of a given trial as to the complete system of events*. For instance, in the experiment with a cube the complete system of events is made up of the events consisting in drawing the numerals 1, 2, 3, 4, 5, 6. The second property states:

The sum of the probabilities of the events forming the complete system equals unity. By virtue of the first property, this sum equals the probability for one of the events forming the complete system to occur. But by the definition of the complete system, one of the events necessarily does take place. (In the example with a cube this means that one numeral out of the six possible numerals has to turn up.) An event that cannot but occur is a certain event, and its probability equals unity. This proves the second property. (In the case of a cube it states that the sum of the probabilities for the six possible numerals equals unity.)

As far as the probabilities of different values that can be assumed by a random variable are concerned, this property was formulated as formula (7) of Chapter 1.

In some cases the reservation citing the incompatibility of events may prove quite substantial when the rule for the addition of probabilities is applied. Consider the following example.

Example. Five marksmen fire simultaneously at one target. The marksmen have identical skills: each hits the target with the probability of 1/3. What is the probability of at least one marksman hitting the target?

We must find the probability of at least one of the following five events (no matter which) to take place: the target is hit by the first marksman, it is hit by the second one, and so forth. A luring thought is to resort to the rule for the addition of probabilities. This rule states that the probability of one of the marksmen to hit the target equals the sum of the probabilities:

$$P = 1/3 + 1/3 + 1/3 + 1/3 + 1/3 = 5/3$$

The result is blatantly absurd. The sum is greater than unity, and this is nonsense. Where did we make a mistake? Recall that the rule for addition is formulated only for incompatible events. Is it excluded that several marksmen hit the target simultaneously? Of course, they can. This is a typical example of compatible events. Hence, the rule for addition is inadmissible here.

To solve the marksmen problem, we have to make use of the rule for the multiplication of probabilities formulated in the next section.

Multiplication of Probabilities

We again throw a cube. The question that has to be answered is the following. We throw the cube twice and obtain two numbers. What is the probability for these numbers to be 6 and 4, and precisely in this order: first 6 and then 4?

The approach to the solution is standard. We make Q trials (each consisting of two throws) and determine the number of trials giving the desired result. First we select all those trials in which the first throw gave 6 regardless of the result of the second trial. This problem is already familiar. All the faces of the cube being identical, the number 6 (or any other number from 1 to 6) appears on the first throw in 1/6 of the trials, that is, at the first stage we select $Q_1 = Q/6$ trials. (The number Q is assumed to be quite large, so that random deviations from Q_1 are small.) Now we want to select those trials in which the second throw gave 4. Each number appears on the second throw again with equal probability. Consequently, the number 4 was found in 1/6 of the trials. Hence, the number of trials in which 4 followed 6 is $Q_2 = 1/6 \cdot 1/6 \cdot Q$, and the probability of this event is

$$Q_2/Q = 1/6 \cdot 1/6 = 1/36$$

Now we introduce complications. Let a trial consist of three throws, and let us find the probability of the trial to yield three numbers in a predetermined order, for instance, 4, 5, 1 or 6, 6, 6. Arguing in a similar manner, we shall find that the number of trials giving the sought result is $Q_3 = 1/6 \cdot Q_2 = 1/6 \cdot 1/6 \cdot 1/6 \cdot Q$, and the probability of this result is

$$Q_3/Q = 1/6 \cdot 1/6 \cdot 1/6 = 1/216$$

Let us analyze another example. Assume that one bicycle per every ten thousand bicycles manufactured by a bicycle plant has an unsound front-wheel axle, and two have unsound rear-wheel axles. Therefore, the probability for a bicycle taken at random to have an unsound front-wheel axle is 1/10 000, and that for the rear-wheel axle is 2/10 000. Assume now that the front- and rear-wheel axles are manufactured in different workshops, so that a defect in one of them does not increase or diminish the probability of a defect in the other. We want to find the probability for a bicycle picked at random to have two unsound axles. We have to argue as in the preceding cases. Among Q bicycles, select those with unsound front-wheel axle. Their number is $Q/10\,000$. Among these select the bicycle with unsound rear-wheel axle. This gives $(Q/10\,000) \cdot (2/10\,000)$. The sought probability is $(1/10\,000) \cdot (2/10\,000) = 2 \cdot 10^{-8}$.

In both examples we fixed the probabilities of several events and wanted to find the probability for these events to occur *jointly*, that is, in the same trial. The results obtained can be formulated in a general statement:

The probability for several events to occur jointly is equal to the product of the probabilities of these events.

This rule must be supplemented with an important clarification. All the above examples operated with *independent events*. Two events are said to be independent *if the probability for one of the events to occur is not affected by whether another event has occurred or not*. For instance, the fact that the first throw of the cube gave 6 does not change in any way the probability to get 4 on the second throw, just as a defect in the front-wheel axle does not change the probability of a defect in the rear-wheel axle.

It is not difficult to understand that the independence of events is essential for deriving the multiplication rule for probabilities.

Let us consider again the example with bicycles and assume that the independence of events is distorted in the following manner. The front- and rear-wheel axles of each bicycle are assembled simultaneously, with the probability of defective axles being higher on certain days than on others. Then the presence of a defect in one of the axles increases the probability of the second axle being unsound, because it increases the probability of the bicycle as a whole to be assembled on an unlucky date. Hence, the probability of both axles being unsound increases.

In order to better understand this point, consider an extreme situation: assume that *all* defective axles are assembled on certain days. All bicycles manufactured on these days have defective rear-wheel axles, and half of them additionally have defective front-wheel axles.

Then the probability for a bicycle selected randomly from the total annual output of the plant to have two defective axles equals the probability of the front-wheel axle to be defective, that is, equals 1/10 000 and not $2 \cdot 10^{-8}$. The multiplication rule for probabilities is thus valid only for independent events.

The multiplication rule yields a straightforward solution of the problem of five marksmen formulated in the preceding section. Let us recall it: five marksmen shoot simultaneously at a target, and the probability for each of them to hit the target is 1/3. What is probability of at least one marksman hitting the target?

The solution is most easily obtained if we calculate the probability for all five marksmen missing (let us denote this probability by P_0). Since the missing or hitting the target by each marksman must be regarded as independent events, the probability P_0 equals the product of the probabilities for each marksman to miss. The event consisting in a marksman hitting the target and the event consisting in this marksman missing it form a complete system of events. The sum of the probabilities of these two events is unity. If the probability for a marksman to hit the target equals 1/3, the probability for him to miss it is $1 - 1/3 = 2/3$. The probability for all five marksmen to miss is

$$P_0 = 2/3 \cdot 2/3 \cdot 2/3 \cdot 2/3 \cdot 2/3 = (2/3)^5$$

The event consisting in all five marksmen missing and the event consisting in at least one of them hitting the target form a complete system of events. The sum of the probabilities of these events equals unity. Therefore, the sought probability P satisfies the equation

$$P + P_0 = 1$$

whence

$$P = 1 - P_0 = 1 - (2/3)^5 \approx 0.87$$

Exercises

1. $\mathscr{N}'$ sites of a network consisting of $\mathscr{N}$ sites are blocked. What is the probability of a randomly selected site to be blocked? nonblocked?

2. Find the probability for three consecutive throws of a cube to give (a) 1, 2, 3 in any order; (b) 1, 2, 2 in an arbitrary order.

3. Workshop A manufactures high-quality parts at a probability 0.8, and workshop B manufactures them at a probability 0.9. Three parts made by workshop A and four parts made by workshop B were taken at random. Find the probability for all seven parts to be good.

Percolation Threshold in a 2×2 Network

The excerpts from probability theory given above are quite adequate for an analysis of a percolation problem in a square network consisting of four sites ($\mathscr{N} = 4$).

Figure 2 gives the schematic of the experiment with a 2×2 network. The figures numbering the four sites are written on separate slips of paper, the slips are put in a hat, and the contents of the hat are well shaken. Assume now that the first extracted piece of paper has a numeral *1*, and the site *1* is blocked (Fig. 2*b*). (The line of argument and the final results do not change in the least if the first to be blocked is a site with a different number. The point is that all sites in a four-sited network occupy equivalent positions.) If the next site to be blocked is site *2*, the current will not be cut (Fig. 2*c*): it will flow through the lower wire. After the third site is blocked (*3* or *4*), the current is certainly interrupted, and one should state that the critical fraction of nonblocked sites is 1/4. If, however, the second to be blocked is site *3* or *4*, the current is interrupted, and the critical fraction

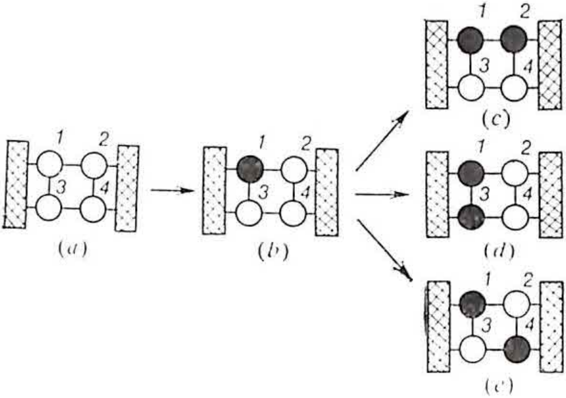


Fig. 2. Calculations for a 2 × 2 network:
(*a*) an initial network; (*b*) one network site is blocked; (*c*), (*d*) and (*e*) two sites are blocked. In case (*c*), current is interrupted only after a third site has been blocked, so that $x_c = 1/4$. In cases (*d*) and (*e*), current is interrupted after a second site has been blocked, so that $x_c = 1/2$. The three cases (*c*), (*d*) and (*e*) are equally probable.

is found to be 1/2 (Fig. 2*d*, *e*). The percolation threshold x_c is therefore a discrete random variable assuming the values 1/4 and 1/2. Let us calculate the probability P for this variable to assume each of these values: P (1/4) and P (1/2).

The decisive factor is which site will be blocked after the first. If it is site *2*, then $x_c = 1/4$, and if it is site *3* or *4*, then $x_c = 1/2$. Therefore, the probability P (1/4) equals the probability of site *2* being the second, and P (1/2) equals the probability of site *3* or *4* being the second. After site *1* has been blocked, all three remaining sites have equal probabilities of being blocked at the next step. The sum of the three probabilities equals unity because these three events form a complete system. Hence, each of these probabilities equals 1/3.

The probability for site *2* to be blocked next is thus 1/3. But if site *2* is the next, then $x_c = 1/4$. Hence, the probability of $x_c = 1/4$ is 1/3, that is, P (1/4) = 1/3. Now we want to find the probability for either site *3* or *4* to be blocked at the second step. By virtue of the addition rule for probabilities, it equals the sum of probabilities: 1/3 + 1/3 = 2/3. This is precisely the probability for x_c to take on the value 1/2; therefore, P (1/2) = 2/3. We must have

$$P\,(1/2) + P\,(1/4) = 1$$

because only two values of x_c are possible. Indeed,

$$P\,(1/2) + P\,(1/4) = 2/3 + 1/3 = 1$$

The mean value of the percolation threshold x_c (4) is readily predictable. According to formula (5) of Chapter 1,

$$x_c\,(4) = 1/2 \cdot P\,(1/2) + 1/4 \cdot P\,(1/4)$$
$$= 1/2 \cdot 2/3 + 1/4 \cdot 1/3 = 5/12$$

This figure differs quite significantly from the threshold value $x_c = \lim\limits_{\mathcal{N} \to \infty} x_c\,(\mathcal{N})$ that was found to be, as we have mentioned already, roughly 0.59.

There are no difficulties in calculating the variance of percolation threshold. According to formula (6) of Chapter 1,

$$\delta^2\,(4) = (1/2 - 5/12)^2 \cdot 2/3 + (1/4 - 5/12)^2 \cdot 1/3$$
$$= 1/72$$

The root-mean-square deviation is

$$\delta\,(4) = \sqrt{2}/12$$

Exercise

4. Retrace the argument, assuming that the first site to be blocked was site *3*.

Continuous Random Variables

So far we have been discussing only discrete random variables. However, there are also *continuous random variables* that can assume any value within a certain interval of the numerical axis.

Assume that a random variable a is allowed to take on any value y in the range from A to B $(A \leqslant y \leqslant B)$, but some of these values occur frequently, while others only rarely. In order to describe this in terms of mathematics, the *distribution function* $f(y)$ of a random variable a is introduced.

The main property of distribution function is as follows: if the points A_1 and B_1 fall within the interval (A, B) and $A_1 < B_1$, *the probability for the value of the random variable to be within the interval* $A_1 \leqslant y \leqslant B_1$ *is equal to the area encompassed by the graph of the function* $f(y)$, *the abscissa axis, and the perpendiculars erected at points* A_1 *and* B_1 (this area is shaded in Fig. 3). The reader familiar with integral calculus will recognize that this probability (let us denote it by $P(A_1, B_1)$) is given by the formula

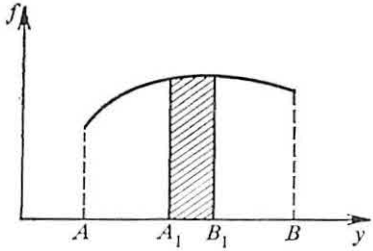


Fig. 3.

$$P(A_1, B_1) = \int_{A_1}^{B_1} f(y)\,dy$$

Since all the values of the random variable are within the interval (A, B) and one of these values is necessarily assumed by the variable, the sought area equals unity. In other words,

$$P(A,\ B) = \int_{A}^{B} f(y)\,dy = 1 \qquad (1)$$

Sometimes this equality is called the condition of normalization of distribution function.

A figure whose area is given by the integral (1) is called a curvilinear trapezoid (see Fig. 3). If the interval $(A_1,\ B_1)$ is so small that the distribution function within this interval remains practically unchanged, the curvilinear trapezoid can be successfully replaced by a rectangle with height $f(y_1)$, where y_1 is an arbitrary point within the interval $(A_1,\ B_1)$. Then

$$P(A_1,\ B_1) = f(y_1)\,\Delta \qquad (2)$$

where $\Delta = B_1 - A_1$ is the width of the interval $(A_1,\ B_1)$.

In the mathematical literature the function $f(y)$ is referred to as the *probability density*. As we see from formula (2), if the interval is sufficiently narrow (otherwise this formula is not valid at all!), the probability for the random variable to fall within the interval is directly proportional to the interval width. The function $f(y)$ is the probability divided by the interval width, or it is the probability per unit length of the interval, or in other words, it is the probability density. Nevertheless, physicists often prefer the term "distribution function".

Formulas (5) and (6) of Chapter 1 for the mean value and variance are rewritten for a continuous random variable in the form

$$\overline{a}=\int_{A}^{B} yf(y)\,dy \tag{3}$$

$$\delta^2=\int_{A}^{B} (y-\overline{a})^2 f(y)\,dy \tag{4}$$

where $\overline{a}$ is the mean value of the continuous random variable a.

Let us consider an example of distribution function.

Uniform distribution. A continuous random variable assumes all values from zero to unity with equal probabilities and cannot assume any other value. Obviously, within the interval (0, 1) the function $f(y)$ is independent of y and vanishes outside this interval (Fig. 4). Its value within the interval is easily obtained from the condition of normalization (1). In this case $A=0$, $B=1$, and the curvilinear trapezoid is transformed into a rectangle with area $f_0\cdot 1$, where f_0 is the value of the function within the interval, and the

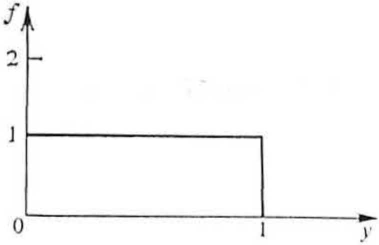


Fig. 4.

interval width is 1. The normalization condition dictates that $f_0 \cdot 1 = 1$, that is, $f_0 = 1$. Therefore,

$$f(y) = \begin{cases} 1 \text{ at } 0 \leqslant y \leqslant 1 \\ 0 \text{ at } y > 1 \text{ and } y < 0 \end{cases} \qquad (5)$$

Exercise

5. A continuous random variable a assumes with equal probability all values from -1 to $+1$. Find the probability for it to fall within the interval from $-3/4$ to $-1/4$.

Percolation Threshold as a Continuous Random Variable

Strictly speaking, percolation threshold is a discrete random variable because all the values that it can assume convert to integers after being multiplied by the total number $\mathscr{N}$ of sites. However, if $\mathscr{N}$ is very large, the difference between the nearest allowed values of this random variable is very small (it equals $\mathscr{N}^{-1}$). Therefore, in this most important case of a very large number of sites, the percolation threshold x_c can be considered with high accuracy as a continuous random variable that takes on all possible values within a certain interval of the numerical axis. Then the quantity x_c must be characterized by the distribution function $f(y)$. In this section we describe the shape of $f(y)$ at large values of $\mathscr{N}$.

The distribution function for threshold values x_c must depend on the number of network sites, $\mathscr{N}$, with which we experiment. Therefore, it is more correct to denote the distribution function by $f_{\mathcal{N}'}(y)$. It will be more convenient to define y not as the threshold value as such, but as its deviation from the mean value $x_c(\mathcal{N}')$. Then $f_{\mathcal{N}'}(y)\,\Delta$ is the probability for the threshold value found in a specific experiment to deviate from the mean value $x_c(\mathcal{N}')$ by a quantity that falls within a small interval Δ in the neighborhood of y. By definition, the mean value calculated by using the function $f_{\mathcal{N}'}(y)$ and formula (3) equals zero.

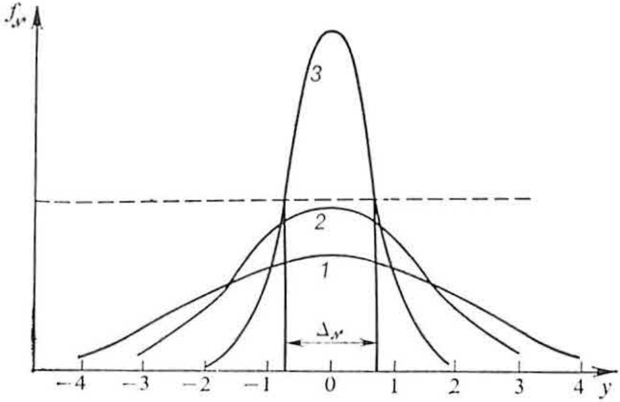


Fig. 5. Functions $f_{\mathcal{N}'}(y)$. The number of sites, $\mathcal{N}'$, increases with increasing curve number. The quantity $\Delta_{\mathcal{N}'}$ is half-width of curve *3* at half-height marked by the dashed line.

Figure 5 represents the function $f_{\mathcal{N}'}(y)$ for three different values of $\mathcal{N}'$. As we see from the graph, the distribution function becomes sharper as the number of sites, $\mathcal{N}'$, increases. This means that as

$\mathcal{N}$ grows, the deviations from the mean value (we remind the reader that it was assumed equal to zero!) become gradually less probable. According to formula (1) of the preceding section, the areas under all bell-shaped curves must be identical. With increasing $\mathcal{N}$ the maximum height of the curves increases, but the width diminishes. The width of a bell-shaped curve can be defined as the distance between the points at which this curve intersects the horizontal line drawn at a distance from the abscissa axis equal to half the maximum height of the curve (see Fig. 5). We denote this width, usually called half-width, by $\Delta_{\mathcal{N}}$.

The probability for the values of percolation threshold to fall beyond the half-width of the curve is lower than that of the most probable value of the threshold by a factor not less than 2. Hence, the half-width characterizes a typical spread of percolation thresholds, including the deviations whose probability is half that at the maximum of the curve $f_{\mathcal{N}}(y)$.

Remember that the root-mean-square deviation carries essentially the same information (Chapter 1). It does not define a deflection whose probability is exactly half as large as the maximum probability, but at the same time characterizes the typical spread of values of percolation thresholds.

The quantities $\Delta_{\mathcal{N}}$ and $\delta_{\mathcal{N}}$ are proportional to each other for any bell-shaped curve, but the proportionality factor depends on the shape of the curve. Computer calculations demonstrate that the percolation threshold distribution function is Gaussian (named after the great mathematician Karl Friedrich Gauss). This function has the following form:

$$f_{\mathcal{N}}(y)=\frac{1}{\delta_{\mathcal{N}}\sqrt{2\pi}}\exp\left(-\frac{y^2}{2\delta^2_{\mathcal{N}}}\right) \qquad (6)$$

where $\exp a \equiv e^a$, and $e \approx 2.72$ is the base of the natural logarithm. It is plotted in Fig. 5 for different values of $\delta_{\mathcal{N}}$. As the function is symmetric with respect to the point $y = 0$ at which it reaches the maximum, the half-width $\Delta_{\mathcal{N}}$ can be found from the relation (see Fig. 5)

$$f_{\mathcal{N}}\left(\frac{\Delta_{\mathcal{N}}}{2}\right)=\frac{1}{2}f_{\mathcal{N}}(0)$$

Making use of formula (6), we obtain

$$\Delta_{\mathcal{N}}=2\,(2\ln 2)^{1/2}\delta_{\mathcal{N}}$$

By virtue of formula (8) of Chapter 1, the quantity $\delta_{\mathcal{N}}$ vanishes as a power function when $\mathcal{N} \to \infty$. This means that as the number of sites grows infinitely, the half-width of percolation threshold distribution function tends to zero, that is, the function itself degenerates to a sharp peak. All values of the percolation threshold, except one, have zero probability. In this connection, we repeat the most important statement of the preceding chapter: as $\mathcal{N} \to \infty$, percolation threshold converts from a random variable into a certain quantity.

Exercise

6. (For readers familiar with integral calculus!) Substitute the function $f_{\mathcal{N}}(y)$ defined by formula (6) into formulas (3) and (4) and prove that the mean value $\overline{a}$ calculated by means of this function is zero, and the variance δ^2 equals $\delta^2_{\mathcal{N}}$.

Chapter 3

Infinite Cluster

This chapter also deals with the site problem of percolation theory, but this time we formulate it in a different language: the language of clusters. Furthermore, we discuss a different object: instead of a network with blocked sites we discuss a doped ferromagnetic, i.e. that with impurity atoms. This is a much more complicated object, and so it must be described at least briefly.

Permanent Magnet

The reason why iron, nickel, cobalt, and some other materials can form permanent magnets is probably known to almost everyone. The explanation of this phenomenon is that the atoms of which such materials are composed are themselves elementary magnets. They possess *magnetic moments*.

The magnetic needle of the compass is a well-known system possessing a magnetic moment. A magnetic moment is a vector. The needle of the compass has southern and northern poles, and its magnetic moment is directed from the southern to the northern pole. The external magnetic field makes the compass needle to turn so that it is oriented along the magnetic lines of force. Of course, any magnetic moment rotates in an external field in the same way. The compass needle produces an external magnetic field.

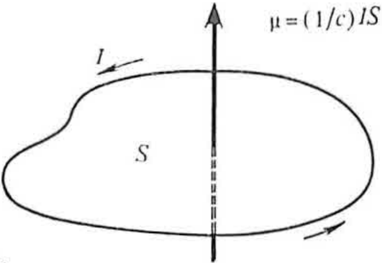


Fig. 6. Current loop and its magnetic moment.

Any magnetic moment produces a completely similar magnetic field.

It was found as early as the beginning of the 19th century that the source of magnetism is the motion of electric charges, that is, the electric current. The magnetic moment is produced by the current. The magnetic moment μ of a planar current loop shown in Fig. 6 is given by the formula

$$\mu = (1/c)\, IS$$

where I is the electric current, S is the loop area, and c is the speed of light (in the CGS system of units). The orientation of the vector is perpendicular to the plane of the loop and such that the current flows counterclockwise if we look in the direction pointed by the vector's arrow.

If the system consists of more than one current loop, we can use the heat-to-tail method for the addition of the magnetic moments of the loops and find the total magnetic moment of the system.

What is the origin of the atomic magnetic moment? Any atom is known to consist of a heavy nucleus and an electron shell. The magnetism of solids originates precisely from the magnetic moment of the shell (the atomic nucleus also can have a magnetic moment, but it is approximately a thousandth of that of the shell).

First, the magnetic moment of the shell grows from the motion of electrons around a heavy nucleus. This rotation can be put in correspondence with a certain current I and an effective area S. Besides, quantum mechanics ascribes to each electron an additional magnetic moment called the *spin moment*. This last moment is in no way related to the characteristics of motion of the electron, but represents its inherent property. However, the spin moment produces a magnetic field just as ordinary moment does. Most often the net magnetic moment of the electron shells of the atoms of which a solid consists is zero. However, in some materials such as iron, nickel, cobalt, and some others the electron shells possess a nonzero magnetic moment.

The magnetic moments of neighboring atoms in a solid interact with one another. In principle, this interaction is similar to the interaction between two compass needles placed close together. Each needle produces a magnetic field acting on the other needle. However, the situation becomes significantly more complicated because the interaction takes place not in the vacuum. The outer electron shells of atoms decisively affect the character of the interaction, up to reversing the directions of the applied forces.

Experiments show that in some materials the interaction between magnetic moments is such

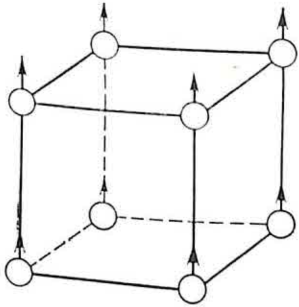


Fig. 7. Fragment of the crystal lattice of a ferromagnetic. The arrows show the orientations of magnetic moments.

that the forces between them make the moments align in the same direction. These materials are called *ferromagnetics* (Fig. 7).

If the magnetic moments of all atoms are oriented in the same direction, the total magnetic moment $\mathcal{M}$ equals the arithmetic sum of individual moments: $\mathcal{M} = \mu\mathcal{N}$, where $\mathcal{N}$ is the number of atoms in the solid, and μ is the magnetic moment per atom.

As the size of a body increases, its magnetic moment grows proportionally to the volume of the body (the number of atoms, $\mathcal{N}$, is proportional to volume). A specific characteristic of magnetic properties, that is, the quantity independent of the size and dependent only on the properties of the atoms composing the body, is the *spontaneous magnetization* M. It is defined as the magnetic moment of a unit volume, that is, it is found as the total moment $\mathcal{M}$ divided by the body volume V:

$$M = \frac{\mathcal{M}}{V} = \mu \frac{\mathcal{N}}{V} = \frac{\mu}{v_0}$$

where $v_0 = V/\mathcal{N}$ is the volume per atom.

The word "spontaneous" emphasizes that the magnetization M is not induced by an external magnetic field but appears because of internal forces. A permanent magnet is just such a body in which the spontaneous magnetization is nonzero. This magnetization produces a magnetic field in the medium surrounding the magnet (or in the vacuum).

The unit of spontaneous magnetization in the CGS system of units is one gauss (1 G). For instance, in iron kept at very low temperatures $M = 1740$ G. From this figure we can find the magnetic moment μ per atom. It is approximately 2.2 of the spin magnetic moment of the electron. The fact that the moment μ is found to be of the order of the spin moment confirms the correctness of our concept of the nature of spontaneous magnetization.

Thermal motion destroys the magnetic ordering, so that there is a critical temperature called Curie temperature (or point), above which the spontaneous magnetization is zero. For instance, the Curie temperature of iron is 770 °C. Iron cannot form permanent magnets at higher temperatures.

Doped Ferromagnetics

Now let us consider a material which is a solid solution (mixture) of magnetic and nonmagnetic (i.e. having a zero magnetic moment) atoms. This is a crystal in whose lattice sites magnetic and nonmagnetic atoms sit, their arrangement being quite random.

Let us assume that the interaction between the magnetic moments of the atoms decreases with distance so fast that we have to take into account only the interaction between nearest neighbors. This means that if two magnetic atoms are at neighboring lattice sites, their moments are necessarily parallel, but if they are separated by at least one nonmagnetic atom, their moments' orientations are arbitrary: they "know nothing" about each other.

The question that we want to pose now is whether the spontaneous magnetization arises in the presence of nonmagnetic atoms, and how many nonmagnetic atoms are needed to destroy the spontaneous magnetization. It will be shown below that the answer to this question reduces to solving the site percolation problem that has been formulated in Chapter 1.

Let us make some definitions. *Two magnetic moments will be said to be connected with each other if they are neighbors or if they are connected via a string of neighboring magnetic atoms* (Fig. 8). The phrase "are neighbors" signifies that the atoms are nearest neighbors. In a square lattice shown in Fig. 8 the nearest neighbors are the neighbors along the horizontal and vertical directions, but not the neighbors along the diagonals. *An ensemble of connected atoms is said to form a cluster.* This definition carries the following meaning. The magnetic interaction makes the connected atoms orient their magnetic moments in the same direction. As a result, each cluster has a resultant magnetic moment proportional to the number of atoms of which it consists. Moreover, we agreed that magnetic atoms which are not nearest neighbors do not interact at all. Consequently, the atoms belonging to different clusters do not interact with

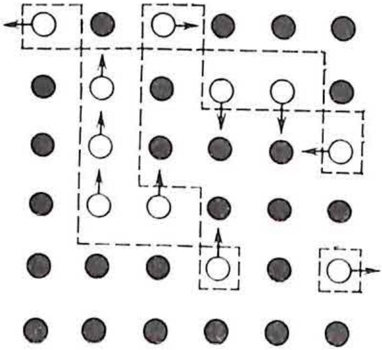


Fig. 8. Fragment of a plane lattice with magnetic (open circles) and nonmagnetic (black circles) atoms. The magnetic atoms form one four-atom cluster, one two-atom cluster, and five one-atom clusters. The dashed lines delineate the cluster boundaries. The moments of different clusters can have different orientations.

one another, and therefore, the relative orientation of the magnetic moments of different clusters is arbitrary (see Fig. 8).

Let us denote by x the fraction of magnetic atoms, that is, the ratio of the number of magnetic atoms to the total number of lattice sites. By definition, x varies within the interval from 0 to 1.

First we assume that the fraction of magnetic atoms is very small ($x \ll 1$). Obviously, in this case they are primarily separated (as raisins in a bun). A cluster made of two magnetic atoms is a rare event, a cluster of three atoms is still rarer, and so on. This statement is extremely important for what follows and will be proved mathematically a little farther below. So far we suggest that those who refuse to recognize this statement as obvious should take it for granted.

At $x \ll 1$ the number of clusters is therefore approximately equal to the number $\mathcal{N}$ of magnetic atoms, and hence, grows proportionally to $\mathcal{N}$ as the total number of lattice sites increases. However, the magnetic moments of these clusters "know nothing" about one another, and hence, are randomly oriented with respect to each other (see Fig. 8). In order to find the net magnetic moment $\mathcal{M}$ of the system, we need to add up the magnetic moments of individual atoms by the head-to-tail method. By virtue of random directions, these moments cancel out, so that the spontaneous magnetization is ultimately zero. *We have thus found that the spontaneous magnetization vanishes at low concentrations of magnetic atoms.*

Formation of an Infinite Cluster

Now consider the case in which almost all atoms are magnetic. Obviously, a small admixture of nonmagnetic atoms does not cancel the spontaneous magnetization but only diminishes it. Let us discuss this effect in terms of clusters. At $x = 1$ all $\mathcal{N}$ atoms belong to a single cluster.

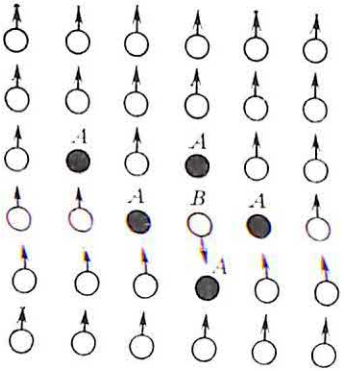


Fig. 9. Fragment of a plane lattice with magnetic (open circles) and nonmagnetic (black circles) atoms at a high concentration of magnetic atoms. With the exception of atom B, all magnetic atoms belong to a single cluster and have identically oriented magnetic moments.

If x only slightly differs from unity, some atoms drop out of this cluster. This occurs, first, because some atoms are replaced by nonmagnetic atoms (atoms A in Fig. 9), and second, because some magnetic atoms form isolated clusters (atom B in Fig. 9) with their own orientation of magnetic moment. Nevertheless, when x is close to unity, a single cluster survives, permeating the whole lattice however large it may be. This *cluster* is said to be *infinite*.

Of course, this concept acquires a rigorous meaning only for an infinite system. Let us take a large series of samples with a given number of magnetic atoms, each containing the same total number of atoms, and in each sample let us find a cluster with the maximum number of magnetic atoms. Now we average the number of magnetic atoms belonging to the maximum cluster over all samples of the series and denote the result of averaging by $\mathcal{N}_{\max}$. Therefore, $\mathcal{N}_{\max}$ is the average number of atoms in the largest cluster. The quantity $\mathcal{N}_{\max}$ is a function of both $\mathcal{N}$ and x. The existence of an infinite cluster has the following corollary: at a given value of x the ratio $\mathcal{N}_{\max}/\mathcal{N}$ tends, when $\mathcal{N}$ increases infinitely, to a nonzero limit

$$\lim_{\mathcal{N}\to\infty} \frac{\mathcal{N}_{\max}}{\mathcal{N}} = P(x)$$

The fraction of atoms $P(x)$ belonging to the largest cluster does not depend on the number $\mathcal{N}$ of atoms if $\mathcal{N}$ is sufficiently large but depends on x. And as $\mathcal{N}$ goes to infinity, the quantity $\mathcal{N}_{\max}$ also tends to infinity. It is for this reason that we speak of an infinite cluster.

Only one infinite cluster can exist in a system. Assume that not only the average number of atoms in the largest cluster is prescribed at the given values of $\mathcal{N}$ and x but also the average number of atoms in the next-largest cluster. Let us denote this last number by $\mathcal{N}'_{\max}$. By definition, $\mathcal{N}'_{\max} < \mathcal{N}_{\max}$. The statement that only one infinite cluster exists in a system signifies that

$$\lim_{\mathcal{N}\to\infty} \frac{\mathcal{N}'_{\max}}{\mathcal{N}} = 0$$

at any value of x. This means that two clusters permeating the system necessarily merge somewhere and transform into a single cluster.*

We have thus obtained that at a sufficiently high concentration x of magnetic atoms, a certain fraction of these atoms belong to one cluster, and they all have a common direction of atomic magnetic moments. This means that the spontane-

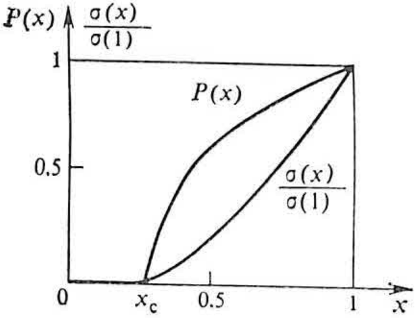


Fig. 10. Graphs of functions $P(x)$ and $\sigma(x)/\sigma(1)$. Although the two functions vanish at a common point, at the critical value x_c they behave quite differently, for reasons to be discussed in Part III of the book.

---

* Strictly speaking, the statement on the survival of a single infinite cluster has not been proved. There are arguments supporting the validity of this conjecture to a greater or lesser extent, but it would be more correct to say that specialists simply take it for granted.

ous magnetization is nonzero:

$$M = \frac{\mu}{v_0} P(x)$$

Recall now that only small clusters exist at a low concentration x of magnetic atoms. In this case an increase in the number of sites, $\mathcal{N}$, results only in the growth of the number of small clusters, but not in a larger number of particles in each cluster. Then

$$\lim_{\mathcal{N} \to \infty} \frac{\mathcal{N}_{\max}}{\mathcal{N}} = 0, \text{ that is, } P(x) = 0$$

We thus have to conclude that there is a critical concentration x_c *at which an infinite cluster is formed,* x_c *satisfying the inequality* $0 < x_c < 1$. At this very concentration x_c the spontaneous magnetization appears, and the function $P(x)$ becomes distinct from zero (Fig. 10). Consequently, *a material cannot form a permanent magnet if the fraction of nonmagnetic atoms is greater than* $1 - x_c$ (the fraction of magnetic atoms is less than x_c).

Exercise

1. Find the function $P(x)$ for x not very different from unity.

Site Percolation Problem Revisited

What remains to be done now is to say that from the point of view of the critical concentration x_c the problem of network conduction and the problem of doped ferromagnetic are identical.

The electric conduction problem can be reformulated just as easily in cluster terms. We then only want to replace the concept "nonmagnetic atom" in all definitions by the term "blocked site".

Figure 8 shows a configuration of magnetic (open circles) and nonmagnetic (black circles) atoms. Let us perform the above-described replacement in this configuration and convert it from a doped ferromagnetic to a wire mesh with cut-out knots. This is done by removing in Fig. 8 the arrows indicating the directions of magnetic moments and then by tracing the wires so that they connect the lattice sites (Fig. 11).

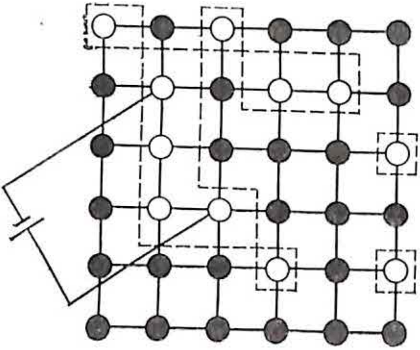


Fig. 11. The same configuration as in Fig. 8, but with nonblocked sites substituted for magnetic atoms.

Figure 11 clearly shows the main property of clusters applied to the network problem. *If a potential difference is applied to any pair of sites within one cluster, a closed circuit with electric current is formed.* (Of course, this property is meaningful only for clusters containing at least two sites.) *When a potential difference is applied to any pair of sites belonging to different clusters, a closed circuit is not formed and electric current is zero.* If $x < x_c$, the system contains only clusters composed of a finite number of sites, so that as the size of the system increases, the current between the lateral electrodes will necessarily be interrupted sooner or later. But if $x > x_c$, a very large system cannot but include into its lateral faces some sites belonging to an infinite cluster. This infinite cluster will create the electric conductivity $\sigma(x)$ distinct from zero and independent of the size of the system.

Let us return to Fig. 10 which shows the functions $P(x)$ (the fraction of the sites belonging to an infinite cluster) and $\sigma(x)/\sigma(1)$ ($\sigma(1)$ is the electric conductivity at $x = 1$, i.e. with no blocked sites). Both functions vanish at the same point that we first identified as the percolation threshold and later the point at which an infinite cluster is born.

We were thus always dealing with a problem of percolation theory that is referred to as the site percolation problem. If we were interested in the value of x_c for a "plane ferromagnetic", then we could look up the result of the experiment with the wire mesh and say that $x_c = 0.59$. However, actual ferromagnetic materials crystallize into three-dimensional, not plane, lattices. An example of a three-dimensional lattice is the primitive cubic lattice whose unit cell has already been shown in Fig. 7.

The problem of wire mesh electric conductance is easily generalized to the three-dimensional case. Imagine a cube consisting of numerous cells and made by welding together pieces of wire, as shown in Fig. 12. We can solder two metal plates on the opposite faces of this cube and, as in Fig. 1, make an electric circuit and study its electric conductance as a function of the number of blocked sites. The blocking of each site interrupts the contact between six wires entering this site. As in the two-dimensional case, there is a critical concentration x_c of nonblocked sites below which the electric conductance vanishes.

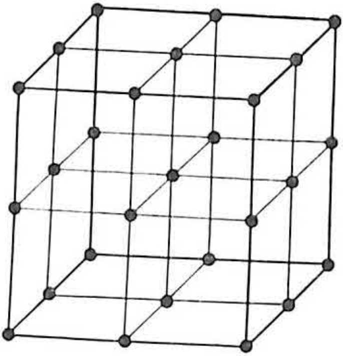


Fig. 12. Simple cubic lattice.

The problem of doped ferromagnetic and the related concept of an infinite cluster were equally valid for plane and three-dimensional lattices. The critical concentration x_c of magnetic atoms at which an infinite cluster appears is at the same time the threshold of percolation from one cube face to the opposite face in a sufficiently large cube. It must be borne in mind that the quantity x_c is strongly dependent on the type of the lattice. It was found to be 0.59 for a square lattice, but $x_c = 0.31$ for a primitive cubic lattice. (For details see Chapter 6.)

Clusters at a Low Concentration of Magnetic Atoms **

The conclusions drawn in the preceding sections were based to a large extent on the statement that at a low concentration x of magnetic atoms these atoms are usually single, two-atom clusters are rare, three-atom clusters are even rarer, and so on. Let us prove this statement.

We introduce a function $P_{\mathscr{M}}(x)$ giving the probability for a randomly chosen atom to belong to a cluster consisting of not less than $\mathscr{M}$ atoms. This means that the atom chosen at random is (a) magnetic, and (b) connected to not less than $\mathscr{M} - 1$ other magnetic atoms. Let us calculate the function $P_{\mathscr{M}}(x)$ for $\mathscr{M} = 1$ and $\mathscr{M} = 2$.

The function $P_1(x)$ is the probability for a randomly chosen atom to be magnetic. This probability equals x (see Exercise 1 to Chapter 2 where the word "nonblocked" must be replaced by "magnetic", and the word "blocked" by "nonmagnetic"):

$$P_1(x) = x \tag{1}$$

The function $P_2(x)$ equals the probability for a randomly chosen atom to be magnetic and to have another magnetic atom among its nearest neighbors. Obviously, these two events are independent, so that the sought probability can be represented by the product of the probabilities of these two events. The first of them (the probability for the atom to be magnetic) equals x, and thus

$$P_2(x) = xW(x) \tag{2}$$

where $W(x)$ is the probability for at least one magnetic atom to be found among the nearest neighbors of the atom. The function $W(x)$ depends on what kind of lattice we consider. Let us limit the analysis to the square lattice

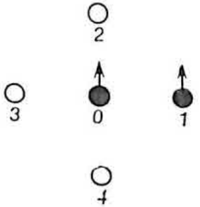


Fig. 13.

in which each atom has four nearest neighbors (Fig. 13). We need to find the probability for at least one of the atoms *1*, *2*, *3*, *4* to be magnetic.

This problem is best solved as follows. The event consisting in all four atoms being *nonmagnetic* and the event consisting in at least one of the four atoms being *magnetic* form a complete system of events. The sum of the probabilities of these two events equals unity. We denote the probability of the first event by W_0; the probability of the second event is just the sought quantity, W. As we have said, $W + W_0 = 1$. The probability for atom *1* to be nonmagnetic equals $1 - x$. The probability for atom *2* or *3* or *4* to be nonmagnetic also equals $1 - x$. The events consisting in different atoms being nonmagnetic are independent. Consequently, the probability for all four atoms to be nonmagnetic equals the *product* of all four probabilities: $W_0 = (1 - x)^4$. This yields $W = 1 - W_0 = 1 - (1 - x)^4$. By virtue of formula (2),

$$P_2(x) = x[1 - (1 - x)^4] \tag{3}$$

If $x \ll 1$, the expression for $P_2(x)$ can be simplified by dropping the terms with high powers of x. Using the binomial formula, we obtain

$$1 - (1 - x)^4 = 4x - 6x^2 + 4x^3 - x^4$$

Note that if $x \ll 1$, the ratio of each next term to its predecessor is small:

$$\frac{6x^2}{4x} = \frac{3}{2}x \ll 1, \ \frac{4x^3}{6x^2} = \frac{2}{3}x \ll 1, \quad \text{and}$$

$$\frac{x^4}{4x^3} = \frac{x}{4} \ll 1$$

Therefore, if $x \ll 1$, we can write to a good accuracy that

$$1 - (1 - x)^4 \approx 4x$$

whence

$$P_2(x) \approx 4x^2 \tag{4}$$

A comparison of formulas (1) and (4) shows that at $x \ll 1$ the ratio

$$\frac{P_2(x)}{P_1(x)} \approx 4x \ll 1 \tag{5}$$

that is, the probability that a randomly chosen atom belongs to a cluster of two or more atoms is much less than the probability for it to form a one-atom cluster.

A simpler derivation of formula (4) for $P_2(x)$ can be given, immediately taking into account the condition $x \ll 1$. It is relatively simple with this derivation to calculate the functions $P_{\mathcal{M}}(x)$ for $\mathcal{M} > 2$ (see Exercise 3). The derivation is as follows: a cluster consisting of more than two atoms, one of them being atom *0*, necessarily incorporates either atom *1* or *2* or *3* or *4*. The probability for atoms *0* and *1* to belong to one cluster equals the probability of both of them being magnetic, and is given by the product of the probabilities for each of these atoms to be magnetic, that is, $x \cdot x = x^2$. The same can be said about the probability for a cluster to be formed by atoms *02*, *03* or *04*. All these probabilities equal x^2. The probability for at least one of these events to realize equals the sum of the probabilities, that is, $4x^2$, hence we arrive at formula (4).

This conclusion holds only for $x \ll 1$. Only under this condition can we use the rule for the addition of probabilities. Indeed, the addition rule holds only for incompatible events. But the event consisting in atoms *0* and *1* being magnetic is *compatible* with the event of atoms *0* and *2* being magnetic. The simultaneous realization of these events means that all three atoms *0*, *1*, and *2* are magnetic, and hence, form a *three-atom* cluster. The probability of the compound event equals the product of the probabilities for all three atoms to be magnetic, that is, equals $x \cdot x \cdot x = x^3$.

If $x \ll 1$, this probability is much less than the calculated probability of the formation of a two-atom cluster. Therefore, the probability for the events to occur jointly can be neglected, and the events can be treated as incompatible. This justifies the derivation given above, under the condition $x \ll 1$.

In fact this means that if $x \ll 1$, *the probability of the formation of a three-atom cluster can be neglected in calculating* $P_2(x)$.

Therefore, in the case of $x \ll 1$ the function $P_2(x)$ actually coincides with the probability for a site chosen at random to belong to a cluster consisting of *two* (and not more than two!) atoms. Correspondingly, the function $P_3(x)$ describes a three-atom cluster. It is proportional to x^3 and is small in comparison with $P_2(x)$. The general result is that the function $P_{\mathcal{M}}(x)$ contains the powers of x not smaller than $x^{\mathcal{M}}$, and that for $x \ll 1$ we should have $P_{\mathcal{M}}(x) \ll P_{\mathcal{M}-1}(x)$.

We thus found that if a site chosen at random in the case of $x \ll 1$ is magnetic, it almost certainly will form a one-site cluster. The probability for it to belong to an $\mathcal{M}$-site cluster sharply drops with increasing $\mathcal{M}$.

Exercises

2. Find $P_2(x)$ for the primitive cubic lattice shown in Fig. 12. Find it for an arbitrary lattice in which each atom has z nearest neighbors.

3. Find $P_3(x)$ for a square lattice, making use of the condition $x \ll 1$.

4. Find $P_3(x)$ for a square lattice, without employing the condition $x \ll 1$.

Chapter 4

Solution of the Site Percolation Problem by Monte Carlo Computer Techniques

The Monte Carlo method is one of the most widespread methods of solving percolation theory problems. The aim of this chapter is to give a general idea of this method, to explain in detail how the main element of the method—the generator of random numbers—operates, and to give, in conclusion, a concrete computer program that makes it possible to determine the percolation threshold of the site problem. The very first question that arises now is:

Why Monte Carlo?

"But what is zéro? You see that croupier, the curly-headed one, the chief one, showed zéro now? And why did he scoop up everything that was on the table? Such a heap, he took it all for himself. What is the meaning of it?"

"Zéro, Granny, means that the bank wins all. If the little ball falls on zéro, everything on the table goes to the bank...."

"You don't say so! And I shall get nothing?"

"No, Granny, if before this you had staked on zéro you would have got thirty-five times what you staked."

"What! Thirty-five times, and does it often turn up? Why don't they stake on it, the fools?"

"There are thirty-six chances against it, Granny."

"What nonsense. Potapitch! Potapitch! Stay, I've money with me—here."

She took out of her pocket a tightly packed purse, and picked out of it a friedrich d'or.

"Stake it on the zéro at once."

"Granny, zéro has only just turned up," I said, "so now it won't turn up for a long time. You will lose a great deal, wait a little, anyway."

"Oh, nonsense; put it down!"

"As you please, but it may not turn up again till the evening. You may go on staking thousands; it has happened."

"Oh, nonsense, nonsense. If you are afraid of the wolf you shouldn't go into the forest. What? Have I lost? Stake again!"*

This excerpt from Dostoevsky's "Gambler" describes the most exciting game of chance of the last century: the roulette. It should be noted that from the standpoint of probability theory, the inexperienced ecstatic grandmother reveals more common sense than the Gambler who is her consultant. The probability of getting zéro will not diminish in the least if it turned up on the preceding round. There is no sense in waiting, as the Gambler advises. This frequently encountered misconception is presumably based on the misunderstanding of the fact that the probability of getting zéro two times running is small. But this does not mean at all that if zéro turned up once, the probability to get it the second time on the next throw is less than on the first. Of course, the probability remains absolutely the same. The city of Monte Carlo in the Monaco principality earned its fame as the world capital of roulette. It was this city that gave its name to one of the most powerful among the modern numerical methods in mathematics.

---

* **"Gambler", in *The Short Novels of Dostoevsky*, Dial Press, 1945, New York; translated from the Russian by Constance Garnett.**

And what is it that this method and the roulette share? It is the fact that the main element of the Monte Carlo method is that very revolving wheel that decides the fates of people, destroying some and rewarding others in numerous casinos of Monte Carlo. Actually, mathematicians have greatly improved it. There is no revolving wheel, but a standard computer program that is called the "random-number generator". But this does not alter the principal point. From the mathematical viewpoint the wheel of the roulette game is nothing but a random-number generator.

What Is the Monte Carlo Method?

As a rule, the term "Monte Carlo method" is applied to any mathematical technique essentially based on a random-number generator.

Usually a modern computer has a standard program that generates random numbers distributed randomly within the interval from zero to unity, that is, it "plays out" the values of a continuous random variable that assumes with equal probability all values within the interval (0, 1).

Each time the program is addressed, it outputs one such number with a predetermined number of decimal places that depends on the computer model.

The simplest use of the Monte Carlo method is for calculating integrals. Imagine, for instance, that we need to calculate the volume within a closed surface of a complicated shape. Let us choose a cube that we know for certain to include the whole surface (Fig. 14). Now we obtain from the random-number generator a set of points distributed uniformly within the cube. This is done as follows. Assume that the cube edge is L long, and all the three coordinates of the points within the cube vary from zero to L (see Fig. 14). Addressing three times the random-number generator, we obtain three numbers y_1, y_2, y_3 within the interval (0, 1). These numbers will give us the coordinates of the first point within the cube by the formulas $X_1 = Ly_1$, $Y_1 = Ly_2$, $Z_1 = Ly_3$. Repeating this procedure Q times, we obtain Q points that on the average uniformly fill the cube. Let Q_1 be the number of points that hap-

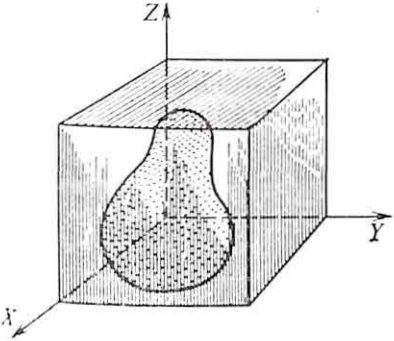


Fig. 14. To the Monte Carlo determination of a pear's volume.

pened to fall within the surface. The points being distributed uniformly, the number Q_1 characterizes the volume enclosed by the surface. Namely, if Q is sufficiently large, the sought volume equals L^3Q_1/Q.

A theory is available that makes it possible to predict the number of points, Q, required to obtain a result with a prescribed accuracy. An empirical approach can be used: we repeat the experiment several times, each time using a new set of random numbers, and then compare the results. If the results fluctuate within the prescribed accuracy, everything is fine and the answer is correct. The Monte Carlo technique has very important advantages over conventional integration methods in the case of multidimensional spaces (multiple integrals).

There are cases in which the Monte Carlo method is the only one practicable. Imagine that we want to study the behavior of a system consisting of an enormous number of particles, for example, the behavior of a gas. In principle, this problem must be solved by the methods of statistical physics; however, these methods are found to be ineffective if the interaction between particles is strong (this happens when density is high and temperature low). Then the properties of the gas are studied by computer simulation. The number of gas particles participating in the simulation is determined by the size of computer memory. The memory must store the information on the coordinates of all particles. The simulation consists in randomly picking up one of the particles which then travels a random distance. (This means that the coordinates of this particle in the computer memory change.) Then another particle is chosen randomly, and so on. The potential energy of the interaction between the particles in the gas depends on their mutual arrangement. The energy is calculated from the very beginning and then is recalculated after each displacement. The probabilities for particles to be displaced by a specific distance are chosen to agree with the potential energy in such a way that the model system "live" on the average as the real system does.

As a result, the computer memory stores what resembles "instantaneous photographs" of the gas taken at successive time moments. The photographs include the coordinates of all gas particles and thus make possible the calculation of mean thermodynamic characteristics, such as pressure, heat capacity, and some others.

The very procedure of simulation much resembles a game which is ruled by a rigid code that includes calling the roulette, that is, the random-number generator. Slight deviations from the rules, or a "crooked" roulette, make some configurations of atoms in the gas appear more often than others; this affects the result of averaging and invalidates the answer.

Random-number generators are used not only in the Monte Carlo method but also in the so-called analogue experiments exemplified by the experiment with a wire mesh described in Chapter 1. We have mentioned earlier that the random sequence of blocked sites required for this experiment was generated on a computer. In order to choose the next site, it is necessary to address the program and obtain a random number y.

We must multiply it by the total number $\mathcal{N}$ of sites and then add unity to the product. Then we have to take the integral part of $\mathcal{N} y + 1$. This gives an integer that falls in the prescribed interval from 1 to $\mathcal{N}$. In fact, such numbers may appear more than once. But there is no harm in it. If it is found that the site with the number just generated was blocked earlier, it is necessary to address the computer for a new random number and convert it into the number of the site.

Somewhat later we shall describe a program with which we can calculate percolation threshold via the Monte Carlo method, but now we shall discuss the principal element of this method, namely, the random-number generator.

How to Think Up a Random Number

We thus need random numbers distributed uniformly in the range from zero to unity. So far the problem is not yet rigorously formulated. We need to know how many decimal places are required in each number. Let us assume that only two decimal places are necessary. Then take a telephone directory, open it at an arbitrary page, and copy a sequence of the last two figures of each telephone number, preceding these numerals with "0.". This will give an acceptable two-digit table of random numbers. And what are we to do if ten-digit random numbers are required? Looks like a computer is then unavoidable.

Come to think of it, the very idea that a computer can generate random numbers may appear strange. Indeed, any computer follows the algorithm that was fed into it, that is, it performs precisely the actions programmed by a man. How then can a chance element be introduced into its performance?

In fact, *there is no chance element in the program of the random-number generator*. The principle of its functioning is as follows. When the program is addressed for the first time, some number y_0 must be fixed. Then a rigorously defined sequence of operations transforms this number into a new number

$$y_1 = \Phi(y_0) \tag{1}$$

where Φ is a specially chosen function or a sequence of operations that transform y_0 into y_1. It is this function that determines the algorithm of random-number generation. In its turn, the number y_1 is the starting point for generating the next number y_2 by the same recipe:

$$y_2 = \Phi(y_1) \tag{2}$$

Obviously, the function Φ is such that all numbers $y_1, y_2, \ldots, y_n$ satisfy the inequalities $0 \leqslant y_n \leqslant 1$. This sequence of numbers represents the sought sequence of random numbers.

It can be readily ascertained that the thus generated sequence cannot be infinite. Indeed, any computer operates only with numbers containing a limited number of decimal places. Their number is limited. (There are only 10^2 two-digit numbers and 10^n n-digit numbers.) Therefore, sooner or later a number y_n will coincide with an earlier generated number, say, y_{n-L}. Then the sequence will be repeated: y_{n+1} will coincide with y_{n-L+1}, and so on.

Consequently, the sequence generated by formulas (1) *and* (2) *is inevitably periodic.* In view of this fact, such numbers are not truly random and are referred to as *pseudorandom* (i.e. "as if random", or resembling random numbers).

Nevertheless, they can be used as random numbers provided the sequence of numbers required to solve a given problem is shorter than the period L of the sequence.

In its turn, the period L is determined by the number of decimal places with which the computer operates (i.e. the number of memory cells allotted to each number) and by the quality of the algorithm (i.e. by the properties of the function Φ that enters into formulas (1) and (2)).

The development of a good random-number generator is a very difficult problem. As a rule, "off-the-cuff" generators prove to be of poor quality. Some specific generators are discussed below.

The Mid-Square Method

Historically, this was the first ever method of random-number generation by a computer. It was suggested in 1946 by the brilliant mathematician John von Neumann. The method makes it possible to generate random numbers with an arbitrary number of decimal places, corresponding to the capabilities of the computer. The method is extremely simple. Assume that we need four-digit numbers. Let us choose the first number X_0 arbitrarily. For instance, $X_0 = 8219$. Raise it to the second power. This gives an eight-digit number $67\underline{5519}61$. Extract the four middle digits: 5519. The second number of the sequence is $X_1 = 5519$. Now raise 5519 to the second power, obtaining $30\underline{\mathbf{4593}}61$. The third random number is then $X_2 = 4593$. If the first digits among the middle group are zeros, the resultant number has a smaller number of nonzero digits. For instance, $X_2^2 = 21\underline{\mathbf{0956}}49$, $X_3 = 956$. Squaring X_3^2, we can obtain an eight-digit number by adding zeros on the left, $X_3^2 = 00\underline{\mathbf{9139}}36$, so that $X_4 = 9139$, and so forth.

Random numbers y_n distributed uniformly within the interval from zero to unity are obtained from the numbers X_n via the formula $y_n = X_n/10^4$, where $n = 0, 1, 2, 3, \ldots$, so that $y_0 = 0.8219$, $y_1 = 0.5519$, $y_2 = 0.4593$, and so on.

At first glance, the method looks attractive. However, a careful investigation demonstrated that this is not true. The main shortcoming of the method is that some starting numbers make the sequence "go into a cycle". For instance, it was found that in the four-digit class of numbers the sequences often terminate by a cycle 6100, 2100, 4100, 8100, 6100. The period of this cycle is mere 4, and this is obviously unacceptable.

There is even a number that immediately reproduces itself. This is 3792 ($3792^2 = 14\underline{\mathbf{3792}}64$). Zero also reproduces itself, and quite often sequences generated by the mid-square method degenerate to zero. Consequently, nowadays the mid-square method is only of historical interest.

Exercises

1. Compose a sequence of four-digit numbers, starting with 0085, 0067, 0032. Show that all three sequences are monotone decreasing (each subsequent number is smaller than the preceding number) and rather rapidly degenerate to zero.

2. Now prove that this constitutes the general shortcoming of the mid-square method: if $2n$-digit numbers X_i are used and a number b appearing in the sequence has zeros for n leading digits, the sequence becomes monotone decreasing and finally degenerates to zero.

Linear Congruent Method

At the present moment this method of generating random numbers is regarded as the best. The idea is as follows. Four integers are chosen:

(1) multiplier k,
(2) shift c,
(3) modulus m,
(4) the first number of the sequence X_0.

The sequence of random numbers is determined by the formula

$$X_{n+1} = (kX_n + c) \bmod m \qquad (3)$$

where the subscript n runs through 0, 1, 2, . . . The symbol $b \bmod m$ denotes the remainder after dividing b by m. For instance,

b	25	6	30	3	147
m	10	10	10	12	12
$b \bmod m$	5	6	0	3	3

Obviously, $b \bmod m < m$. Consequently, all numbers of the sequence, X_n, satisfy the inequality $X_n < m$. The sequence of the numbers y_n distributed uniformly within the interval from zero to unity is obtained by the formula

$$y_n = \frac{X_n}{m}, \quad n = 0, 1, 2, \ldots \tag{4}$$

It is not just any choice of the four starting numbers that leads to good results. Note first of all that the sequence X_n must necessarily be periodic, and the period cannot be greater than m. Indeed, all X_n being integers, with $X_n < m$, the number of different numbers cannot exceed m. For this reason, a number that has already occurred in the sequence will appear at least beginning with $n = m$, and the sequence will repeat itself.

However, it is far from simple to generate a sequence with a maximum possible period $L = m$. If the starting numbers are not carefully selected, the generated sequences will have, as a rule, short periods.

Exercises

3. Write the sequence of numbers X_n generated by means of formula (3) for $k = 3, c = 0, X_0 = 5, m = 20$.

4. Write the sequence of numbers X_n generated by means of formula (3) for $k = 3, c = 1, X_0 = 5, m = 20$.

5. Write the sequence of numbers X_n generated by means of formula (3) for $k = 3, c = 2, X_0 = 5,$

$m = 20$. Make sure that in all three cases the period of the sequences is essentially shorter than 20. Analyze other examples.

The following **theorem** is true. *If a sequence is generated by means of formula (3) for* $c \neq 0$, *its period equals* m *if and only if the following conditions are met:*

(i) c *and* m *are coprime numbers (have no other common divisor than* 1);

(ii) $b = k - 1$ *is a multiple of* p *for any prime* p *which is a divisor of* m;

(iii) b *is a multiple of* 4 *if* m *is a multiple of* 4.

Unfortunately, the proof of this theorem is too complicated to be given here.

6. Make sure that the conditions imposed by the theorem given above did not hold in all the examples presented in Exercises 3-5.

7. Make sure that the set of integers $k = 11$, $c = 3$, $m = 5$ meets the conditions of the above theorem and yields the period $L = 5$ for arbitrary X_0.

Therefore, a generator with a maximum possible period L will be obtained if we take for m the largest number with which a given computer can operate, and choose the other numbers in accordance with the theorem given above.

However, the period is not the only characteristic of the quality of a random sequence. For instance, let us consider a sequence generated with $k = c = 1$. The sequence is $0, 1, 2, 3, \ldots, m - 1, 0, 1, 2, 3, \ldots, m - 1, 0, \ldots$ Its period is indeed m, but as a random sequence it is absolutely unacceptable.

A complicated system of tests has been elaborated to determine the quality of a random-number generator. Therefore, only reliable generators can be advised for applications.

When a random-number generator is being chosen, the properties of the computer are important not only from the standpoint of the choice of the period of maximum possible length. The rate at which random numbers are generated also depends on the choice of the starting numbers. And it is found that different generators prove to be optimal for different types of computers.

Programs involving Monte Carlo computations often have to address a random-number generator an enormous number of times (tens and hundreds of millions of times). Consequently, high speed is one of the most important characteristics of a generator.

The generator recommended for a BESM-6 computer has $k = 5^{17}$, $c = 0$, $m = 2^{40}$, and odd X_0. This set of numbers does not meet the conditions of the above theorem ($c = 0$), and the period of this generator is less than m. However, another theorem has been proved for generators with $c = 0$, and by virtue of this theorem, the period of the recommended generator is $2^{38} \approx 2.75 \cdot 10^{11}$.

Determination of Percolation Threshold by Monte Carlo Simulation on a Computer. Distribution of Blocked and Nonblocked Sites

Now we shall describe in detail a computer program that determines the percolation threshold by Monte Carlo techniques. Note that this program is not unique. In fact, each group of researchers that deals with these problems prefers to work with their own program that has some individual features. This stems from specific features of different computers and to some extent from the experience accumulated by individual programmers.

The problem that we mean is the site problem, and for the sake of simplicity, we consider only a two-dimensional square lattice. Actually, it will become clear later that a generalization of the method to an arbitrary lattice of arbitrary dimensionality is readily obtainable.

Let us look at percolation in a square with the side containing L sites, so that the total number of sites is $\mathcal{N} = L^2$. We assume the distance between the sites to be unity, and describe the sites by their coordinates X and Y. For instance, a site with coordinates $X = 9$, $Y = 25$ is the site in the ninth column from the left and in the twenty-fifth row from below.

In order to study percolation, we must fix which of the sites are blocked and which are not, and we must be able to vary the number of blocked sites in order to exceed the percolation threshold. To achieve this, we first assign a specific number V to each site. Each site being characterized by two coordinates X and Y, this is equivalent to introducing a function of two variables $V(X, Y)$ whose arguments X and Y do not run through all possible values but are allowed to assume only integral values within the interval from 1 to L. Programmers call such a function a two-dimensional array, and the values taken on by this function are said to be the elements of this array. For instance, the element $V(31, 97)$

of an array is a number assigned to the site with coordinates $X = 31$, $Y = 97$. Altogether the array V has $L \times L = \mathscr{N}$ elements, and we have to reserve in the computer memory a space necessary to store $\mathscr{N}$ numbers.

The work of the program begins with generating this array. Its elements are random numbers uniformly distributed between zero and unity. A random-number generator outputs a number y, and this number is assigned to the array element $V(1, 1)$. This means that this number is written into the corresponding memory cell of the computer, and that from this moment on the computer "memorizes" that $V(1, 1) = y$. The next number output by the generator is assigned to the element $V(1, 2)$, and so forth. Thus all the elements of the array V are generated.

Then a second two-dimensional array, that we denote by K, is formed. The elements of this array are zeros and unities, so that if, for example, $K(25, 16) = 0$, this means that the site with coordinates $X = 25$, $Y = 16$ is blocked, but if $K(25, 16) = 1$, this site is nonblocked. The array K is generated by using the array V and a certain number t that falls within the interval from zero to unity. Varying t, we can change the number of blocked sites.

The array K is generated by obeying the following rule. Choose a site with coordinates X and Y. If $V(X, Y) \leqslant t$, then $K(X, Y) = 1$, and if $V(X, Y) > t$, then $K(X, Y) = 0$. The site with coordinates X and Y is treated as nonblocked in the former case and as blocked in the latter case. The quantities V being uniformly distributed within the interval from zero to unity, we can assume t being close to zero and obtain that almost all sites are blocked. Conversely, if t is close to unity, almost all sites are nonblocked. When $t = 1/2$, the number of blocked and nonblocked sites must be nearly equal.

Making use of the distribution of random numbers output by the generator, it is possible to relate the quantity t to the mean fraction x of nonblocked sites obtained as a result of the above-described procedure. It can be shown (see Exercise 8) that $t = x$. However, this equality is valid if we take a very large number $\mathcal{N}$ of sites or if we generate many arrays K with the same t and then average the fractions of nonblocked sites found in each array. Actually, in each concrete array, x may deviate from t to some extent, in both directions, but the equality holds the better, the larger $\mathcal{N}$ is.

The computer thus stores the array V, from which we can generate an array K that describes which site is blocked and which is not. The form of the array K is dictated by the quantity t which is approximately equal to the fraction of nonblocked sites generated in this array. Smoothly varying t, we can generate the distributions of blocked and nonblocked sites with smoothly varying concentration of nonblocked sites, x.

Exercise

8. Prove that the mean fraction x of nonblocked sites in an array K equals t.

Search for Percolation Path

Let us assume that the array V is stored, the criterion t is fixed, and the array K with a certain fraction of nonblocked sites is found.

Now the computer knows precisely which site is blocked and which is not, and we pass to the second stage of the program, namely, the search for percolation path. Assume that we look fo percolation from left to right. First we replace all unities in the leftmost column ($X = 1$) by twos. The replacement consists in erasing a unity in the memory cell corresponding to a given element of the array K and in writing a two instead. A list of coordinates of the sites equipped with twos is compiled in the computer memory. Then the computer analyzes each site of this list. The computer calculates the nearest neighbors of the site under study and requests from the array K the information concerning these neighbors. If the nearest neighbor contains a unity, it is given a two, and its coordinates are added to the new list. After the first list has been analyzed, the computer memory contains a list of "second-generation" twos, that is, the list of unities replaced by twos because they were in contact with the first-generation twos.

In order to economize on the computer memory, the first list is erased at this stage: it is no longer necessary; the appropriate memory cells are emptied. The computer starts an analysis of the second list and the formation of the third-generation twos. When this has been completed, the second list is erased and an analysis of the third list is initiated. This analysis generates the fourth list, and so on.

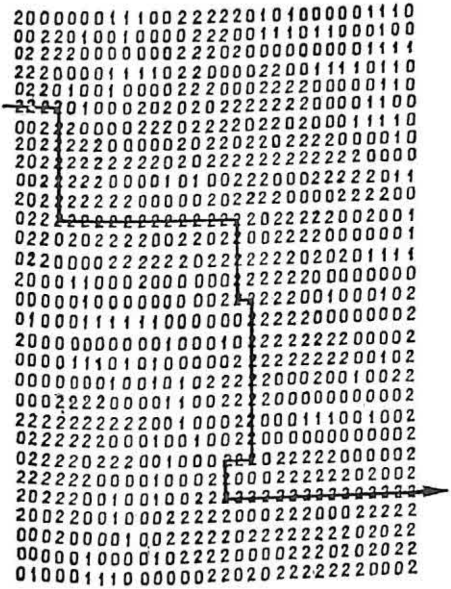
The number of twos in the array K grows in the course of this procedure. Twos are nonblocked sites linked by percolation to a nonblocked site of the leftmost column, that is, the twos label the percolation path.

The search for percolation path terminates in two cases:

1. A two appears on the right-hand side of a square. The computer recognizes the existence of percolation at the given value of t.
2. There are no twos on the right-hand side of a square, and the analysis of the last list has not generated new sites marked with 2. This means that all paths broke down, and there is no percolation at this t.

Determination of the Threshold

Assume that the computer recognized percolation at a given t. Then it diminishes t, and, making use of the same array V, finds a new array K with a reduced number of nonblocked sites. The search for percolation paths is then repeated. If percolation is found again, t is further reduced, until at a certain t percolation ceases. Then the interval between this value of t and the minimum value at which percolation was found is divided in two, and percolation paths are looked for at this intermediate value of t. If it is found that percolation paths are cut, the interval between this last value and the minimum value at which percolation was detected is again divided in two. If percolation is found, the interval to be divid-



```
200000011100222220101000001110
002201001000022200111011000100
022220000000222020000000001111
222000011110220000220011110110
022010010000222000222200000110
222201000202020222222200001100
002220000222022220220200011110
202222200000202202202222000010
202222222220202222222222220000
002222200001010022200022222011
202222222000002022220000222200
022222222222222222022222002001
022020222200220220022220000001
022000022222202222222020201111
200011000200000022222200000000
000001000000000222222001000102
010001111110000002222200000002
200000000001000102222222200002
000011101010000022222222200102
000000010010102222200020010022
000222200001100222200000000002
222222222200100022000111001002
022222200010010022000000000002
022220222001000222022222000002
222222000010002200022222202002
202220010010022222222222222222
200220010002222200022200022222
000200001002222202222222202022
000001000010222200002220202022
010001110000002202022222220002
```

Fig. 15. Pattern of the distribution of zeros, unities, and twos at the moment of percolation onset. The broken line shows the path by which twos "percolated" from the left-hand side to the right-hand side of the square. In this case the computer did not stop calculations when the first "2" appeared on the right-hand side, but continued until new twos ceased to appear.

ed in two is the one between the last value of t and the value at which no percolation was found.

Percolation threshold is thus "bracketed" with an arbitrary precision. If we find no percolation at the first chosen value of t, then t must be increased until percolation is detected, and this must be followed by "bracketing". This method enables us to find t, corresponding to the percolation threshold, with any prescribed accuracy. At this t the computer finds the fraction of non-blocked sites, x, which is close to the value of t but is not necessarily equal to it. This value of x is taken for the percolation threshold found in this experiment. Figure 15 shows the first percolation path from left to right found in a 30 × 30 square.

Identical experiments are then rerun many times with different sets of random numbers in the array V. This corresponds to changing the random sequence of blocked sites in the experiments with a wire mesh. The results of these experiments make it possible to determine the mean value $x_c(\mathscr{N})$ of percolation threshold for a prescribed number of sites, $\mathscr{N}$. (For this we have to add up all the obtained values of percolation threshold and divide the sum by the number of runs.)

In order to find the true percolation threshold $x_c = \lim\limits_{N \to \infty} x_c(\mathscr{N})$, it is necessary to vary the number of sites, $\mathscr{N}$, in the square and determine the function $x_c(\mathscr{N})$. This function must then be approximated by an analytical expression of the type*

$$x_c(\mathscr{N}) = x_c(\infty) + \frac{D}{\mathscr{N}^{\gamma}} \qquad (5)$$

that is, we have to choose three quantities, $x_c(\infty)$, D, and γ such that expression (5) fit in the best way the results obtained by means of the computer. If this fit is achieved so that $\gamma > 0$, we can say that it is the quantity $x_c(\infty)$ that gives us the limiting value of x_c. Indeed, by virtue of expression (5),

---

* There is no rigorous answer to the question why this expression contains $\mathscr{N}$ in the power-law manner, but the results of many numerical experiments show that this is invariably the case in all percolation theory problems.

$$\lim_{\mathcal{N} \to \infty} x_c(\mathcal{N}) = x_c(\infty)$$

The accuracy achieved by this procedure is the better, the greater the amount of data necessary to establish the dependence of $x_c(\mathcal{N})$. In its turn, this amount of data is limited by the speed and memory of the computer employed.

Exercise

9. Look carefully at Fig. 15 and reconstruct the way by which individual groups of twos were formed.

Part II

Various Problems of Percolation Theory and Their Applications

Chapter 5

Problems on Two-Dimensional Lattices

We Are Planting an Orchard (the Bond Problem)

Imagine that a vast orchard is being designed. The fruit trees in the orchard must be planted not arbitrarily but in a regular manner. They are to be located at the sites of a periodic lattice drawn on the ground. Many such lattices can be invented, but the following three will be sufficient for us here: a square, a triangular, and a hexagonal lattices (the latter is often referred to as the "honeycomb lattice"). The lattices are shown in Fig. 16. Land being expensive, it is natural to try and plant the trees as close to one another as possible, but this cannot be done for a number of reasons. One of the reasons is that the designers are afraid of infectious diseases of the trees. Let us assume that experts on tree infections supplied the following information*:

---

* The author is not responsible for this information, and thus prays the reader not to be too serious about practical conclusions drawn from the solution of the problem formulated above. The problem is given merely to illustrate the potential inherent to percolation theory.

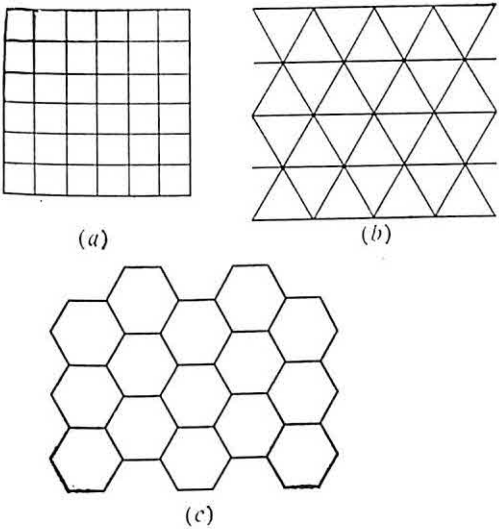


Fig. 16. Plane lattices: (*a*) square; (*b*) triangular; (*c*) honeycomb.

1. A tree blighted by the disease can infect only its nearest neighbors.

2. Some nearest-neighbor trees inevitably infect one another when one of the trees has been infected. In other cases the disease is not transferred (e.g. this may depend on the separation between the branches of a given pair of trees). A pair of trees in which the infection is inevitable will be referred to as a *connected pair*.

3. The experts also supply us with the form of the function $x(a)$, mainly, the probability for a pair of nearest neighbors chosen at random to be connected. This probability depends on the distance a between the nearest neighbors in a

given lattice. Obviously, the function $x(a)$ increases with decreasing argument a: two trees infect each other the easier, the smaller the separation between the trees.

We are to answer the following question: What is the number of trees that a diseased tree can infect? This question can be given only a probabilistic answer. If a given tree forms a connected pair with one of its neighbors, this neighbor is infected. In their turn, the blighted trees infect their neighbors, and so forth. Therefore we can only ask: What is the probability for an infected tree to transfer the disease to a concrete number of trees in the orchard?

At this juncture it is convenient to switch to the cluster language introduced in the preceding chapter. We assume that two neighbor sites with two trees forming a connected pair are linked by a *bond* represented by a piece of wire connecting the two sites. If two nearest-neighbor trees do not form a connected pair, the bond between them (the wire) is broken (Fig. 17).

Two sites will be regarded as *connected* if they are linked by an unbroken bond, or if they are linked by an unbroken chain of sites that are nearest neighbors and are linked by unbroken bonds (e.g. the sites A and B, as well as the sites C and D, in Fig. 17 are connected).

An ensemble of connected sites will be said to form a cluster. In the context of the given problem, the most important property of a cluster is that a diseased tree infects all the trees of its cluster and none outside this cluster.

By definition, the fraction of unbroken bonds equals x. Further argument runs as in Chapter 3.

When x is small, unbroken bonds are mostly single, almost all clusters consist of two sites, three-site clusters are infrequent, and four-site clusters are even less so. When x is large, there exists an infinite cluster of connected sites. When $x = 1$, this cluster comprises all the sites in the system. As x diminishes, this cluster loses some of the sites, and finally, at a certain critical value x_c, the infinite cluster ceases to exist.

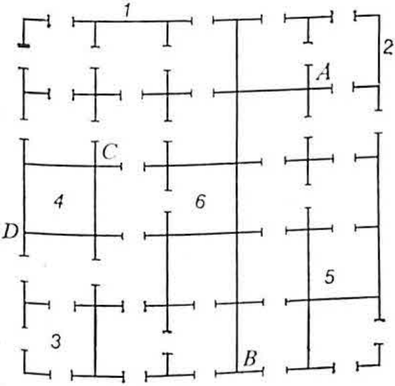


Fig. 17. Fragment of a square lattice with broken bonds. Shown in the figure are three two-atom clusters (*1*, *2*, *3*), one four-atom cluster (*4*), one six-atom cluster (*5*), and one ten-atom cluster (*6*).

The infinite cluster is precisely the catastrophe from which the orchard must be saved. Let $P^b(x)$ be the probability for a randomly chosen site to belong to the infinite cluster. If $x < x_c$, so that $P^b(x) = 0$, one blighted tree can infect only several other trees. And if $x > x_c$, one blighted tree infects an infinite number of trees in an infinite orchard with a probability $P^{\mathrm{b}}(x)$. Therefore, if $x < x_{\mathrm{c}}$, the pocket of disease accidentally brought to the orchard remains localized in the vicinity of the infected site, but it propagates all over the orchard if $x > x_{\mathrm{c}}$.

In order to arrive at practical recommendations, it is necessary to find x_{c} and equate it to the function $x(a)$ supplied by the experts. This yields the separation a_{c} as the solution to the equation $x(a_{\mathrm{c}}) = x_{\mathrm{c}}$. The source of infection in the orchard remains localized if the separation between the trees exceeds a_{c}, otherwise the infection propagates throughout the orchard.

Percolation theory refers to the above-enunciated problem of determining x_{c} as the *bond problem* in order to underline that here the random element lies in the *bonds* that can be unbroken or broken with a given probability. At first glance, the bond problem is quite similar to the site problem discussed in the preceding chapters. However, the two problems are not reducible to each other on a given lattice and have different answers.

We shall have to somewhat complicate our notations in this and subsequent chapters. Let us denote the percolation threshold of the site problem by x_{s}, and that of the bond problem by x_{b}. These thresholds depend on the type of the lattice. Let us use the following abbreviations for the types of two-dimensional lattices: S for square, T for triangular, and H for hexagonal, or honeycomb. Then $x_{\mathrm{s}}(\mathrm{H})$ denotes the percolation threshold of the site problem on a hexagonal lattice, $x_{\mathrm{b}}(\mathrm{T})$ the percolation threshold of the bond problem on a triangular lattice, and so on.

The function $P^{\mathrm{b}}(x)$ introduced in this section for the bond problem should not be confused with the function $P(x)$ defined earlier for the site problem.

The bond problem can be formulated not only in cluster terms but also as a problem of percolation from one side of a square to the opposite side. Recall the experiment with a wire mesh with which we started the book. Some readers may have wanted to ask why it was necessary to cut simultaneously the four wires entering each site, instead of cutting randomly chosen individual wires (bonds). Now it can be clearly understood that by cutting the bonds the researchers would find $x_{\mathrm{b}}(\mathrm{S})$ instead of $x_{\mathrm{s}}(\mathrm{S})$ which was actually determined in their experiment. Now we can explain why the site problem has been initially selected: it will be shown below that on a square lattice the bond problem has an exact analytical solution which yields that $x_{\mathrm{b}}(\mathrm{S}) = 0.5$. Consequently, it would not be reasonable to carry out such a time- and labor-consuming experiment for the sake of $x_{\mathrm{b}}(\mathrm{S})$, with $x_{\mathrm{s}}(\mathrm{S})$ being known only from approximate solutions.

Exercise

1. Find the function $P^{\mathrm{b}}(x)$ for $1 - x \ll 1$ for the three lattices shown in Fig. 16.

Inequality Relating x_{b} to x_{s}

When analyzing the site problem, the function $P(x)$ is often replaced by a function $P^{\mathrm{s}}(x)$

related to $P(x)$ by a relation

$$P(x) = xP^{\mathrm{s}}(x) \tag{1}$$

By definition, $P(x)$ is the probability for a randomly chosen site to belong to an infinite cluster. It can be written as the product of the probabilities of two independent events. In the language of the ferromagnetics problem, the first of these events is for a randomly selected site to be magnetic. The probability of this event equals x (see Exercise 1 to Chapter 1). The second event consists in this site being connected to an infinite cluster of magnetic sites. Therefore, the function $P^{\mathrm{s}}(x)$ defined by formula (1) is the probability for a randomly selected magnetic site to be connected to an infinite cluster. In other words, $P^{\mathrm{s}}(x)$ is the fraction of *magnetic* sites belonging to the infinite cluster, that is, the ratio of the number of sites belonging to the infinite cluster to the number of magnetic sites. We remind the reader that $P(x)$ is the ratio of the number of sites belonging to an infinite cluster to the total number of sites. Naturally, the function $P^{\mathrm{s}}(x)$ grows monotonically with increasing x, equals unity for $x = 1$, and vanishes for $x \leqslant x_{\mathrm{s}}$.

The English mathematician Hammersley, who was the first to enunciate percolation theory, proved a theorem that states that

$$P^{\mathrm{s}}(x) \leqslant P^{\mathrm{b}}(x) \tag{2}$$

Both functions P^{s} and P^{b} increase monotonically with increasing argument x. Therefore (Fig. 18), formula (2) implies that

$$x_{\mathrm{b}} \leqslant x_{\mathrm{s}} \tag{3}$$

that is, the threshold for the bond problem is not greater than that for the site problem on any lattice (not necessarily two-dimensional). This result can be rewritten as a different inequality:

$$1 - x_b \geqslant 1 - x_s \tag{4}$$

that permits the following interpretation. Let us assume that we need to block electric current

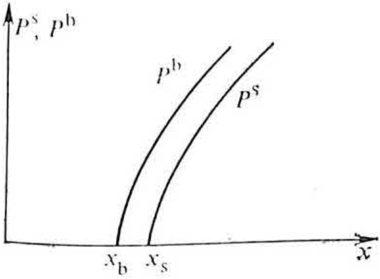


Fig. 18.

through a network of wires or block the flow of liquid through a network of pipes, and this can be done either by blocking network sites or by cutting network bonds (wires or pipes) connecting these sites. Inequality (4) means that the system is more readily blocked by cutting out its sites than by cutting the bonds. The fraction of blocked sites that terminates the flow is less than the fraction of broken bonds giving the same result. This conclusion appears quite natural because not one but all bonds entering a site are cut when this site is blocked.

Exercise

2. Find the function $P^s(x)$ for $1 - x \ll 1$ for the three lattices shown in Fig. 16. Compare the result with that obtained in Exercise 1 and check whether inequality (2) holds.

Clarification. In Exercise 1 to Chapter 3 we recommended that you should find the function $P(x)$ for the site problem, for $1 - x \ll 1$. However, there we meant only the principal term of the function, that is, $P(x) = x$. If this expression is substituted into formula (1), we obtain $P^s = 1$. This is the correct result in the sense that $\lim_{x \to 1} P^s(x) = 1$. We suggest that the reader find the small terms that make the function $P^s(x)$ differ from unity. Obviously, as $x \to 1$, these terms have zero for their limits. Consequently, the result may be written in the form

$$P^s(x) = 1 - A(1 - x)^n$$

where A and n are positive numerical coefficients depending on the type of the lattice.

Covering and Containing Lattices

The site problem is more general than the bond problem. The bond problem is reducible to the site problem but on a different lattice said to cover the former. A *covering lattice* is constructed by using the following procedure:

1. Place a site of the covering lattice in the middle of each bond of the initial lattice.

2. Connect two sites of the covering lattice if and only if the bonds of the initial lattice on which these sites were placed meet at a site of the initial lattice.

The result of such a construction is a new periodic lattice which is said to cover the initial lattice.

Figure 19 shows the covering lattice in the case of an initial square lattice. Thin lines trace the initial square lattice. Semicircles show the places where the initial lattice has sites. The covering lattice consists of thin and solid lines, but there are no sites of the covering lattice where thin lines cross. The sites lie only at the crossings of solid lines and are marked by black circles.

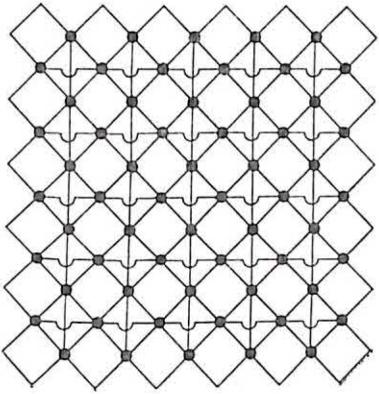


Fig. 19. Covering lattice for the square lattice.

Each bond of the initial lattice joins three other bonds at one end and three more bonds at the other end. Hence, each site of the covering lattice must be connected with six other sites. This is shown in Fig. 19. Each site is connected with four other sites by solid lines, and with two more sites by thin lines.

Now assume that a bond problem has been formulated on the initial lattice, that is, a certain fraction of randomly selected bonds has been cut.

Assume now that if a bond of the initial lattice is broken, the site of the covering lattice placed at this bond is blocked. We thus obtain a site problem on the covering lattice. Its sites are randomly blocked, and the fraction of blocked sites equals the fraction of broken bonds in the initial lattice.

Note that the existence of an infinite cluster of connected sites in the bond problem inevitably signifies the existence of an infinite cluster made of coupled unbroken bonds. Conversely, the absence of an infinite cluster of sites signifies that bonds do not form an infinite cluster.

As follows from the method of construction of the covering lattice, the existence of an infinite cluster formed by unbroken bonds in the initial lattice signifies that there exists an infinite cluster formed by nonblocked sites in the covering lattice, and conversely, if there is no infinite cluster of bonds in the initial lattice, there is no infinite cluster of sites in the covering lattice. Therefore, the percolation threshold of the bond problem in the initial lattice equals the percolation threshold of the site problem in the covering lattice. If the initial lattice is denoted by L, and the covering lattice by L_{cov}, this statement can be written as formula

$$x_{\text{b}}(L) = x_{\text{s}}(L_{\text{cov}}) \tag{5}$$

Introducing the concept of a *lattice containing another lattice*, it is possible to derive a number of inequalities relating percolation thresholds on a number of lattices. Assume, for instance, that lattice L is obtained from lattice L_{cont} by crossing out a certain number of bonds. Then lattice L_{cont} is said to *contain lattice* L.

For instance, let us take a triangular lattice. If we erase all the bonds that are marked in

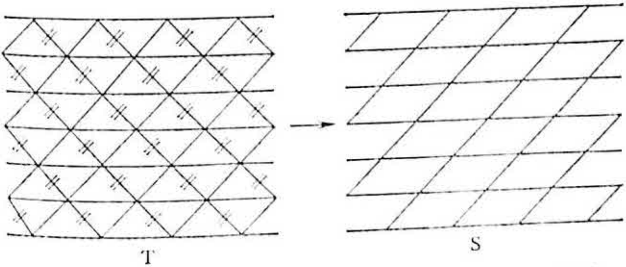


Fig. 20. The triangular lattice contains the square lattice.

Fig. 20 by two short lines, it will transform into the lattice shown on the right-hand side of the drawing. It can be easily seen that from the standpoint of the site or bond problems this new lattice is equivalent to a square lattice. Indeed, the fact that the angles between the bonds of the new lattice are not 90° is found to be immaterial if we analyze the relations between different sites (the new lattice can be just "straightened out"). The percolation threshold of the bond problem (as well as of the site problem!) on this lattice is exactly equal to the percolation threshold on the square lattice. For this reason, a triangular lattice is said to *contain* a square lattice.

Now let us assume that a certain fraction of bonds in the containing lattice is broken. The bonds of the containing lattice can be classified into the bonds that are common for the containing and contained lattices, and the bonds that are specific for the containing lattice (these last are marked by two short lines in Fig. 20). The bonds being broken quite randomly, the fraction of broken bonds in one category of bonds is absolutely the same as in the other category, and equals the fraction of broken bonds in the whole lattice. Therefore, in order to obtain a contained lattice with the same fraction of broken bonds, it is necessary to break additionally those bonds of the containing lattice that are intact but are specific for this lattice, that is, are completely absent in the contained lattice.

This argument demonstrates that the number of unbroken bonds leaving each site of the containing lattice is not less than (is greater than or equal to) the number of unbroken bonds leaving the same site of the contained lattice. As a result, the probability for a randomly selected site to belong to an infinite cluster of the containing lattice is not less than that of the contained lattice. This statement implies the inequality

$$P_L^{\mathrm{b}}(x) \leqslant P_{L_{\mathrm{cont}}}^{\mathrm{b}}(x) \tag{6}$$

The left-hand side of inequality (6) includes the function $P^{\mathrm{b}}(x)$ calculated for the contained lattice, and the right-hand side includes this function for the containing lattice. By analogy to inequality (2) employing inequality (3), inequality (6) implies that

$$x_{\mathrm{b}}(L_{\mathrm{cont}}) \leqslant x_{\mathrm{b}}(L) \tag{7}$$

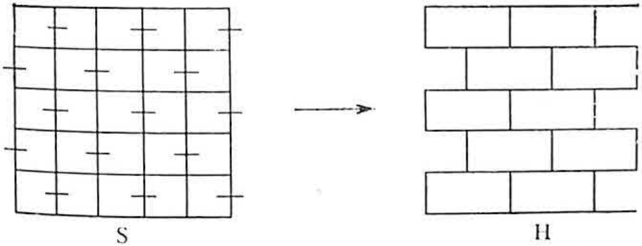


Fig. 21. The square lattice contains the honeycomb lattice.

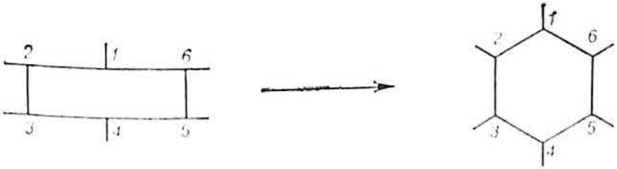


Fig. 22. Transformation of one unit cell of the lattice shown on the right-hand side of Fig. 21 into a honeycomb unit cell.

that is, the containing lattice has a lower x_{b} than the contained lattice.

It has been already mentioned that a triangular lattice contains a square lattice. For this reason,

$$x_{\mathrm{b}}\,(\mathrm{T}) \leqslant x_{\mathrm{b}}\,(\mathrm{S}) \tag{8}$$

Now assume that some bonds are deleted in a square lattice, as shown in Fig. 21. This yields the lattice shown on the right-hand side of the drawing. Look attentively at this drawing. It is equivalent to a hexagonal lattice. Just pull it slightly upwards, somewhat warping the bonds, and its cells (Fig. 22) will transform into the "honeycomb" lattice shown in Fig. 16. Consequently, a square lattice *contains* a hexagonal lattice, and hence,

$$x_b(S) \leqslant x_b(H) \tag{9}$$

Inequalities (8) and (9) imply that

$$x_b(T) \leqslant x_b(H) \tag{10}$$

Let us turn now to the site problem. Assume that the same sites are blocked in the containing and contained lattices (correspondingly, both have equal fractions of blocked sites). Let the contained lattice have an infinite cluster of intact sites. This means that it also exists in the containing lattice because additional bonds can only facilitate its formation. But if it is known that an infinite cluster exists in the containing lattice at a given fraction of sites, no conclusion can be drawn from this as to the existence of an infinite cluster in the contained lattice. The removal of some bonds between sites in the transition from the containing to the contained lattice can be "lethal" for the infinite cluster. Consequently, the percolation threshold of the site problem in the contained lattice cannot be smaller than that in the containing lattice:

$$x_s(L_{\text{cont}}) \leqslant x_s(L) \tag{11}$$

Therefore, we can write a chain of inequalities for the triangular, square, and hexagonal lattices:

$$x_s(T) \leqslant x_s(S) \leqslant x_s(H) \tag{12}$$

quite identical to that for the bond problem.

Let us turn again to Fig. 19 where the covering lattice for the square lattice is shown. Imagine that the bonds traced by thin lines have been deleted. Obviously, this gives us a square lattice merely rotated through 45°, which, of course, is quite immaterial for percolation theory problems.

The covering lattice for a square lattice thus contains a square lattice. By virtue of formula (5),

$$x_{\mathrm{s}}\,(L_{\mathrm{cov}}) = x_{\mathrm{b}}\,(\mathrm{S}) \tag{13}$$

where L_{cov} is interpreted as the covering lattice shown in Fig. 19. However, inequality (11) and the fact that this covering lattice contains a square lattice imply that

$$x_{\mathrm{s}}\,(L_{\mathrm{cov}}) \leqslant x_{\mathrm{s}}\,(\mathrm{S}) \tag{14}$$

Inequalities (13) and (14) imply that

$$x_{\mathrm{b}}\,(\mathrm{S}) \leqslant x_{\mathrm{s}}\,(\mathrm{S}) \tag{15}$$

We have thus obtained that the bond problem threshold on a square lattice is lower than the site problem threshold. Inequality (15) is a particular case of Hammersley's general theorem that is embodied in formula (3) and was given earlier without proof. This theorem was not used in deriving (15), and we thus can say that the arguments given above prove this theorem in the case of a square lattice.

"White" Percolation and "Black" Percolation

Now let us look at the bond problem from a slightly different viewpoint. So far we normally have said that there are intact and broken bonds distributed randomly over a lattice and applied the term "cluster" to an ensemble of sites connected by unbroken bonds.

The problem can be reformulated in a more symmetric manner. Let us rename broken bonds "black", and unbroken bonds "white". An ensemble of sites connected by white bonds will be referred to as a white cluster, and an ensemble of sites connected by black bonds as a black cluster (in our former terminology it was a "white" cluster that we called a cluster). Let the fraction of white bonds be denoted, as before, by x. The fraction of black bonds will be denoted by q. Each bond has to be either black or white, so that $q = 1 - x$.

After this reformulation we can speak of percolation through both white bonds and black bonds.

When the concentration x of white bonds is low, there is no infinite white cluster, but there is an infinite black cluster, that is, an infinite cluster of sites connected through black bonds. Conversely, when the concentration q of black bonds is low (i.e. at x nearly equal to unity), there is an infinite white cluster but not an infinite black cluster.

When x varies from zero to unity, two events take place: the black infinite cluster vanishes, to be replaced by a white cluster, or, to say the same in different terms, percolation through black bonds vanishes, to be replaced by percolation through white bonds. But what is the sequence in which these events take place?

The white and black bonds differ only in labels. It is obvious, therefore, that the critical concentration q_b at which percolation appears through black bonds equals the concentration x_b at which percolation appears through white bonds.

Therefore, as x increases, percolation through white bonds appears when $x = x_b$, and percolation through black bonds stops when $x = 1 - q_b = 1 - x_b$. The sequence in which

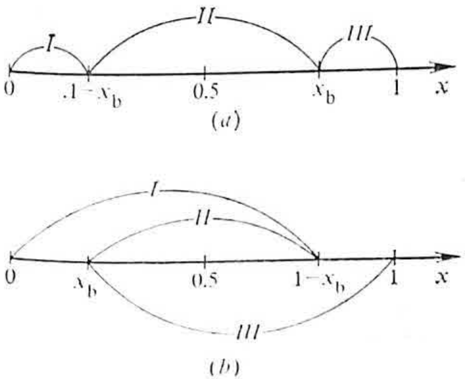


Fig. 23.

these events take place depends on the sign of the difference $x_b - 0.5$.

If it is necessary for percolation through white bonds that the bonds be more than 50% white (this means that the number of white bonds must exceed that of black bonds at the percolation threshold), then at increasing x first percolation through black bonds ceases and then percolation through white bonds appears (Fig. 23*a*). In region *I* in Fig. 23*a* there is percolation only through black bonds, in region *III* only through white bonds, and in region *II* there is no percolation through either white or black bonds.

If $x_b < 0.5$, first percolation through white bonds appears and only then percolation through black bonds ceases. In region *I* in Fig. 23*b* there is percolation through black bonds, in region *III* it proceeds through white bonds, and in region *II* through both black and white bonds.

The same symmetric reformulation can be found for the site problem. We remind the reader that in this case all bonds are unbroken, but sites can be of two sorts. In the wire-mesh problem we called them nonblocked and blocked. In the ferromagnetics problem these were magnetic and nonmagnetic sites. And now we introduce, as in the bond problem, a universal notation: nonblocked, or magnetic, sites are said to be *white*, while blocked, or nonmagnetic, sites are said to be *black*. White sites are said to be connected if they are nearest neighbors or are connected by a string of white sites that are formed by pairs of nearest neighbors. Black sites can be connected in the same symmetric manner.

We are justified in speaking of percolation through white and black sites. If $x_s > 0.5$, there is a region of x values with no percolation through either white or black sites ($1 - x_s < x < x_s$). If $x_s < 0.5$, the region $x_s < x < 1 - x_s$ is the region in which percolation proceeds both through white and black sites.

The symmetric approach proves constructive because sometimes it is possible to conclude from an analysis of the lattice that it cannot accommodate either percolation through white or percolation through black, or that one of these must necessarily be present. This information is sufficient to make readily certain conclusions on percolation threshold.

For instance, let us consider the site problem on a triangular lattice. Assume that the lattice percolates through white sites. You will readily see that this excludes percolation through black sites. Let us assume that we study percolation of current from left to right on a very large network, as in the experiment whose description opened this book. But now the network is not a square but a triangular lattice. It is easy to understand that the structure of any triangular lattice is such that a rightward percolation path through white sites excludes any possibility for downward percolation through black sites. Indeed, black sites cannot "penetrate" through the broken line traversing the entire network from left to right and connecting white sites.

As we explained in Chapter 3, an infinite cluster ensures percolation in any direction if the size of the system is sufficiently large. It then follows that a triangular lattice cannot contain, for the same value of x, two coexisting infinite clusters, one of white and another of black sites, that is, percolation through white sites and percolation through black sites are incompatible. Hence,

$$x_s\,(\mathrm{T}) \geqslant 0.5$$

The same conclusion follows for square lattices:

$$x_s\,(\mathrm{S}) \geqslant 0.5$$

A theorem has been proved for triangular lattices, stating that $x_s\,(\mathrm{T}) = 0.5$. The proof of this theorem cannot be given here, but the main points can be grasped more or less easily, making use of the concepts of percolation through white and percolation through black. Drawing various configurations of black and white sites, it can be noticed that blocked rightward percolation through white sites necessarily *implies* downward percolation through black sites (square lattices do not possess this property!).

A triangular lattice thus cannot simultaneously have percolation through white sites and percolation through black sites, but there must always be percolation through one sort of sites. This means that region *II* in Fig. 23 (x_b in this figure must now be replaced by x_s) degenerates to a point, that is, x_s (T) $= 0.5$.

In the cases of bond problems of this type an analysis can be facilitated by operating with dual lattices.

Dual Lattices

Only plane lattices can be *dual*. By definition, a plane lattice is a lattice that can be placed on a plane, with bonds intersecting only at points where lattice sites are located. For instance, all the lattices shown in Fig. 16 are plane, while the covering lattice in Fig. 19 is not plane because its bonds intersect at points shown by arcs, but there are no sites at these crossing points. (These arcs act as "bridges" that uncouple rightward and downward roads.)

Each plane lattice divides the plane into cells. Lattice L^d is said to be dual to lattice L if each bond of L^d intersects one and only one bond belonging to L, and vice versa, each bond of L intersects one and only one bond belonging to L^d.

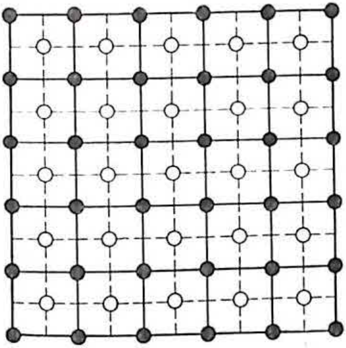


Fig. 24. Construction demonstrating that square lattices are self-dual. The solid lines and black circles show the bonds and sites of the initial lattice; the dashed lines and open circles show the bonds and sites of the dual lattice.

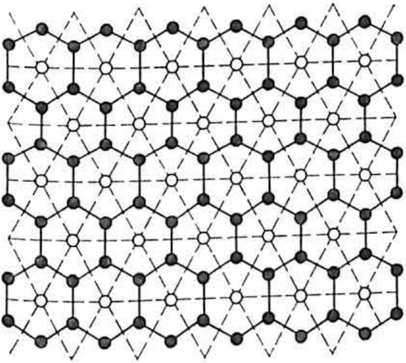


Fig. 25. Construction demonstrating that a triangular lattice is dual to a honeycomb lattice (and vice versa). Notations are the same as in Fig. 24.

Furthermore, each cell of lattice L must contain only one site of lattice L^d (and vice versa).

You see from the definition that duality is a mutual property: if L^d is dual to L, then L is dual to L^d. Figures 24 and 25 show that a square lattice is dual to a square lattice, and a triangular and a hexagonal lattices are mutually dual.

Let us return now to the bond problem. We assume that if a bond in the initial lattice is white (unbroken), the bond of the dual lattice intersecting the former is black (broken). Therefore, if the concentration of white bonds in the initial lattice is x, the concentration of white bonds in the dual lattice is $q = 1 - x$.

In what follows it will be more convenient to interpret percolation threshold as that concentration of white bonds at which conductivity from left to right first appears (or disappears) in a very large wire mesh with electrodes soldered to it (see Fig. 1). This formulation of the problem is identical to that discussed at the very beginning of the book, but now the disrupted elements of the net are the bonds and not the sites of the lattice.

Assume now that a rightward percolation path exists through white bonds of the initial lattice. It is easy to understand that this signifies no downward percolation through white bonds in the dual lattice. Indeed, by definition, a white bond of the initial lattice is intersected only by a black bond of the dual lattice. Hence, if the initial lattice contains a nowhere disconnected broken line composed of white bonds and traversing the entire network from left to right, this means that the white bonds of the dual lattice cannot at any point "break through it" and form a downward path.

If a lattice is sufficiently large, the existence of a rightward path on the initial lattice signifies that the fraction of white bonds exceeds the threshold value:

$$x > x_\mathrm{b}(L) \tag{16}$$

Conversely, the absence of a downward path through white bonds of the dual lattice indicates that the fraction of white bonds, $q = 1 - x$, in the dual lattice is less than the threshold fraction for the dual lattice, that is, $1 - x < x_\mathrm{b}(L^\mathrm{d})$, or

$$x > 1 - x_\mathrm{b}(L^\mathrm{d}) \tag{17}$$

According to all this, all the values of x satisfying inequality (16) also satisfy inequality (17). Hence, $x_\mathrm{b}(L) \geqslant 1 - x_\mathrm{b}(L^\mathrm{d})$, or

$$x_\mathrm{b}(L) + x_\mathrm{b}(L^\mathrm{d}) \geqslant 1 \tag{18}$$

On a square lattice $L = L^\mathrm{d}$, so that inequality (18) implies that

$$x_\mathrm{b}(\mathrm{S}) \geqslant 0.5 \tag{19}$$

and making use of inequality (15), we additionally obtain

$$x_\mathrm{s}(\mathrm{S}) \geqslant 0.5 \tag{20}$$

that is, we come to the same conclusion as in the preceding section. Recall that the experiment with a square lattice described in Chapter 1 gave $x_\mathrm{s}(\mathrm{S}) = 0.59$, which does not contradict inequality (20).

It has been rigorously proved for the square and triangular lattices that inequality sign in formula (18) must be replaced by equality sign, that is,

$$x_{\mathrm{b}}(L) + x_{\mathrm{b}}(L^{\mathrm{d}}) = 1 \tag{21}$$

This immediately leads to two new results:

$$x_{\mathrm{b}}(\mathrm{S}) = 0.5 \tag{22}$$

and

$$x_{\mathrm{b}}(\mathrm{T}) + x_{\mathrm{b}}(\mathrm{H}) = 1 \tag{23}$$

We shall not give here the rigorous proof of formula (21): it would require that we introduce a number of new concepts that will not be necessary in the further exposition. However, this formula can be given a sufficiently clear interpretation. Tracing out various configurations of white and black bonds, it can be discerned that in the case of square and triangular lattices the absence of rightward percolation through white bonds in the initial lattice always signifies that there exists a downward percolation path through white bonds in the dual lattice. Let us take this statement for granted. Assume that the fraction of white bonds is such that there is no rightward percolation on the initial lattice. For a sufficiently large network this means that

$$x < x_{\mathrm{b}}(L) \tag{24}$$

Correspondingly, the dual lattice contains a a downward percolation path. The fraction of white bonds in this lattice is $1 - x$. Therefore, $1 - x > x_{\mathrm{b}}(L^{\mathrm{d}})$, or

$$x < 1 - x_{\mathrm{b}}(L^{\mathrm{d}}) \tag{25}$$

The above arguments signify that any value of x satisfying inequality (24) must also satisfy inequality (25). This is possible if $1 - x_b(L^d) \geqslant x_b(L)$, or

$$x_b(L) + x_b(L^d) \leqslant 1 \tag{26}$$

Inequalities (18) and (26) yield equality (21).

Formula (23) alone is not sufficient for finding separately x_b (T) and x_b (H). However, if we make use of the star-delta transformation known in the theory of electric circuits, we arrive at a relation between percolation thresholds of the bond problem on the triangular and hexagonal lattices. As a result, each of these thresholds is calculated:

$$\begin{aligned} x_b(\mathrm{T}) &= 2\sin(\pi/18) \approx 0.347\,296 \\ x_b(\mathrm{H}) &= 1 - x_b(\mathrm{T}) = 1 - 2\sin(\pi/18) \approx \\ &\approx 0.652\,704 \end{aligned} \tag{27}$$

Exercise

3. Look at Fig. 53 (see p. 252) illustrating the star-delta transformation. If you understand by looking at this drawing how to write the relation mentioned above and arrive at formulas (27), you have done very well. If you do not go that far, do not despair because the problem is not that simple. The derivation of formulas (27) first obtained in 1963 by the English mathematicians Sykes and Essam was an important event in percolation theory. Now read attentively the text placed alongside the figure and you are likely to feel pleased with the beautiful application of probability theory yielding such a nontrivial result.

Results for Plane Lattices

In conclusion we must give a summary table for percolation thresholds for plane lattices (Table 1).

Table 1

Percolation Threshold for Plane Lattices

Lattice type	x_b	x_s
Triangular	0.3473	0.5
Square	0.5	0.59
Hexagonal	0.6527	0.70

Only two figures in this table, namely, x_s (S) and x_s (H), were obtained by approximate methods. All the others represent exact solutions. In the next chapter we will show that the situation with three-dimensional lattices is much less satisfactory. In fact, *not a single* exact solution has been obtained. You need not regard this situation as strange. No universal methods exist for solving the problems of percolation theory analytically. Each exact solution mentioned above looks like a miracle. Rather, we should be surprised how many exact solutions have already been devised.

Exercise

4. Let us revisit the orchard mentioned at the beginning of this chapter. Assume that the distance separating the trees is chosen so as to satisfy the requirement that the fraction of connected pairs equals the threshold value. Let the function $a(x)$ giving the distance between neighboring trees as a function of the fraction of connected pairs, x, be known. Naturally, the greater x is, the shorter a is because the trees infect one another the easier, the closer to one another they are planted. If $x = x_{\mathrm{b}}$, the separation between trees equals $a(x_{\mathrm{b}})$. Find the area per tree for the thus chosen separation, for three different lattices. The lattice with the least area per tree is the most profitable. Could you say in advance which of the lattices gives the least area per tree, starting only with the information that the function $a(x)$ monotonically decreases with increasing x?

Directed Percolation

Assume now that the trees planted at the sites of a plane lattice form not an orchard but a forest, and that this forest catches fire. Some nearest-neighbor trees have intertwined branches and readily pass on the fire. In accordance with our general terminology, we shall say that the sites containing such trees are connected by white bonds. Other nearest-neighbor trees do not light each other up. We shall say that the corresponding sites are connected by black bonds. The white and black bonds are distributed randomly over the lattice, with the fraction of white bonds being equal to x.

Our problem is to find the critical value x_b such that at $x < x_b$ the pocket of fire remains localized, while at $x > x_b$ the fire spreads over the whole forest.

Obviously, this is simply one more example of the bond problem. The value of x_b can be found by looking it up in Table 1 (see p. 116).

Assume now that the forest fire is accompanied by a strong wind, so that the flames propagate only along the wind. This results in a new and interesting problem, called the *problem of directed percolation*.

We postulate that the forest was planted at the sites of a square lattice, and the wind is blowing along the diagonal of the square cells. In Fig. 26 the wind direction is indicated by the arrow at the top of the figure, and the lattice has been rotated through 45° with respect to the standard arrangement of square lattices in this book. The problem is now reformulated as follows. Each white bond is converted into a vector whose arrow is placed so that the projection of the vector onto the wind direction is positive. As before, black bonds do not let the fire pass whatever the direction, while white bonds let it pass only in the direction of the arrow. We want to find the critical fraction of white bonds beginning with which one tree in flames in an infinitely large forest can start a fire spreading to an infinitely large distance.

The solid lines with arrows in Fig. 26 stand for white bonds; two percolation paths are shown: *1* and *2*. Path *1* traces the locus of the fire, while

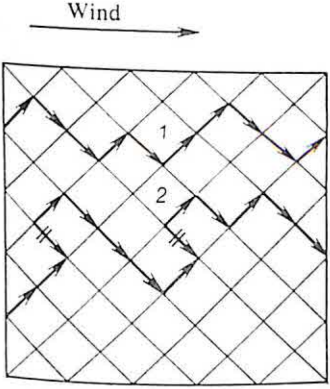


Fig. 26. Directed white bonds are shown by the solid lines with arrows; black bonds are shown by the thin lines. Flame propagates from left to right along path *1*, and cannot propagate along path *2*. Two segments of path *2* on which the fire would have to move against the wind are marked by two dashes.

path *2* does not correspond to the conditions of directed percolation: at two points marked by dashes the propagation goes counter to a white bond, that is, against the wind.

We thus find that while in nondirected percolation white bonds worked in both directions, in directed percolation they work only in one direction. Hence, the critical fraction of white bonds in directed percolation, x_{b}^0, cannot be smaller than in conventional percolation, that is, $x_{\mathrm{b}}^0 \geqslant x_{\mathrm{b}}$.

At the present time, approximate solutions have been obtained for a number of problems with directed percolation. Thus, in the above-described bond problem on a square lattice

$x_b^0 = 0.63$ or 0.64 (the results yielded by several methods slightly differ from one another). Recall that in the nondirected bond problem on a square lattice we had $x_b = 0.5$.

Quite a few physical problems are reducible to directed percolation. As an example, we can cite the motion of electrons in a strong electric field in a randomly inhomogeneous medium. These are media with properties that vary randomly from point to point, so that a moving electron runs into obstacles that it has to bypass. At the same time, the uniform electric field acts as a wind that drives electrons in one direction only.

Directed percolation also realizes in the problem of the electric conductance of a wire net (see Chapter 1) if we assume that each bond between the sites of the net contains a diode (a rectifying element) that lets the current flow in one direction only. Then the fraction of broken bonds at which the current through the net drops to zero corresponds to the threshold of directed percolation. A combined problem, with diodes present not in each bond of the net, has also been analyzed.

Chapter 6

Three-Dimensional Lattices and Approximate Evaluation of Percolation Thresholds

As was shown in Chapter 5, percolation threshold essentially depends on lattice type. The purpose of the present chapter is to explain qualitatively what exactly are the properties of lattices that are important for percolation thresholds. Having understood this aspect, it is possible to predict percolation thresholds (to within 10%) without solving the problem. The ability to come up with such predictions (be the prediction accurate or not) appears to be valuable because, first, the number of different lattices is quite substantial, and second, the computation of a threshold takes (on top of a high skill) about half an hour of computation time of the most advanced computers.

Moreover, percolation problems are not restricted to lattice problems only. We shall be able to see that applications mostly require solving problems that are formulated not on lattices. It is found that the ideas of approximate evaluation presented in this chapter are quite fruitful for nonlattice problems. It was on the basis of these ideas that a number of percolation thresholds for nonlattice problems had been predicted with high accuracy long before these problems were solved with computers.

In order to collect the necessary experience, we have to go beyond the framework of plane lattices discussed in the preceding chapter and turn to three-dimensional lattices.

Three-Dimensional Lattices

The simplest of the three-dimensional (or simply 3D) lattices is the *simple cubic lattice*. It is shown in Fig. 12. Its unit cell is a cube shown in Fig. 27. The vectors $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$ are called *translation vectors*. Elongating each of these translation vectors an integral number of times (n_1, n_2, n_3) and then adding up the resultant vectors, we can obtain a vector $\mathbf{R}_{n_1, n_2, n_3}$ beginning at the origin of coordinates and ending at any site of the simple cubic lattice:

$$\mathbf{R}_{n_1, n_2, n_3} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3$$

In cubic lattices the three vectors $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$ are of identical length a, so that the numbers n_1, n_2, n_3 are simply three Cartesian coordinates of lattice sites in units of a. The simple cubic lattice is said to be generated by *translation* (parallel displacement) of the cubic unit cell by vectors that are integral multiples of $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$.

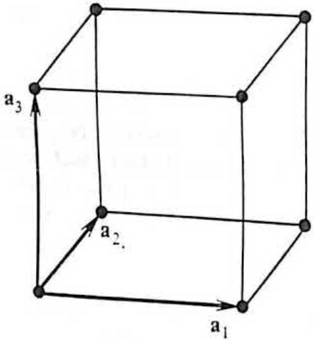


Fig. 27. Unit cell of a simple cubic lattice.

The most important characteristic of a lattice is the number of nearest neighbors (also called the *coordination number*), denoted by z. In the simple cubic lattice $z = 6$.

Such alkali halides as NaCl (common salt), KCl (rock salt), LiF, NaI, and a number of others crystallize into the simple cubic lattice. The alkali metal ions (e.g. Na^+) alternate in this lattice with the halogen ions (e.g. Cl^-).

Body-centered cubic lattice (abbreviated to bcc). This lattice can be composed of two simple

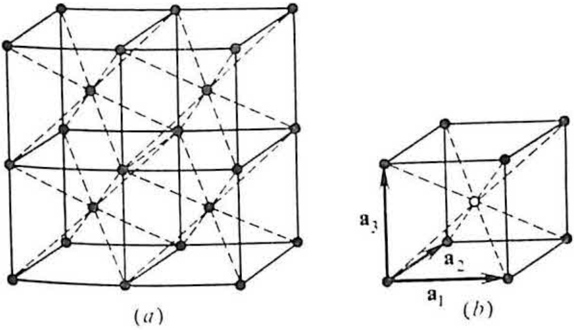


Fig. 28. (*a*) Body-centered cubic lattice; (*b*) unit cell of the bcc lattice.

cubic lattices by displacing them relative to each other by one half of the body diagonal of the unit cell. (By definition, the body diagonal of a cube is the line connecting the opposite vertices of the cube and passing through the central point.)

A bcc lattice is shown in Fig. 28*a*. Its unit cell is shown in Fig. 28*b*. The entire bcc lattice can be generated by translating this unit cell by vectors $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$; however, it will not be sufficient to place the vector $\mathbf{R}_{n_1, n_2, n_3}$ at the origin if we want to generate all sites of the lattice. We also need to place the origin of the vector at the central atom. Only then will the two simple cubic sublattices of the bcc lattice be translated together.

The nearest neighbors of each site of a bcc lattice lie in the direction of the body diagonal of the unit cell (see Fig. 28*b*). Each site has eight such neighbors, so that $z = 8$. The distance to a nearest neighbor equals one half of the body

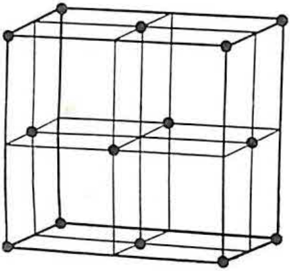


Fig. 29. Unit cell of a face-centered cubic lattice.

diagonal, that is, $\sqrt{3}a/2$, where the letter a denotes, as before, the length of the edge of the unit cell.

Univalent alkali metals, such as lithium, sodium, potassium, rubidium, cesium, bivalent barium, and a number of other substances crystallize into bcc lattices.

Face-centered cubic lattice (abbreviated to fcc). A unit cell of an fcc lattice is shown in Fig. 29. In comparison with a simple cubic lattice, it contains additional sites placed at the centers of each face. In order to translate this cube, the origin of the vector $\mathbf{R}_{n_1, n_2, n_3}$ must be placed at the origin of coordinates and also at the centers of the three nonopposing faces of the cubic unit cell. The nearest neighbors of each site of an fcc lattice lie in the direction of face diagonals. The distance to a nearest neighbor is $a/\sqrt{2}$. Each of the three mutually perpendicular plane intersecting at a given site contains four nearest neighbors of this site, so that the number of nearest neighbors, z, of a site is 12.

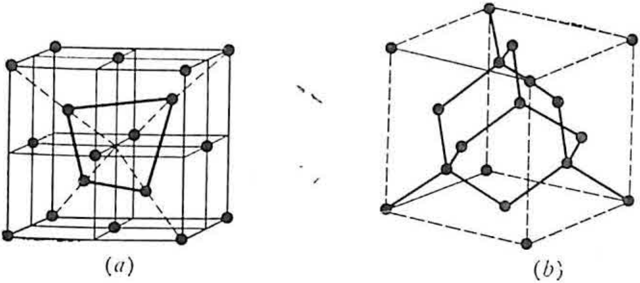


Fig. 30. (*a*) Diamond lattice; (*b*) tetrahedral structure of bonds in the diamond lattice.

Substances that crystallize into fcc lattices are such metals as copper, silver, gold, aluminium, lead.

The last lattice that we introduce into our analysis is the *diamond lattice*. It is shown in Fig. 30*a*. You obtain it if you imagine two face-centered cubic lattices displaced with respect to each other along the body diagonal of the cubic unit cell by one fourth of its length.

Substances that crystallize into diamond-type lattices are elements of the fourth group of the periodic table: carbon (diamond), as well as the two most important semiconductor elements: germanium and silicon. All these elements are tetravalent, and the atoms in their lattices are bonded by covalent forces. To simplify, we can imagine that each atom has four "arms" corresponding to four valence electrons. An atom in a lattice then holds hands with its four nearest neighbors. The diamond lattice is perfectly suited to this type of bonding. Each site of this lattice lies at the center of a regular tetrahedron formed by other sites (Fig. 30*b*). The number of nearest neighbors is $z = 4$.

Percolation Thresholds for 3D Lattices

The bond and site problems are formulated for 3D lattices exactly as for plane lattices. As there, we again assume that bonds connect only nearest-neighbor sites.

Table 2 summarizes percolation thresholds of the site and bond problems in the 3D lattices described above. As we have already mentioned, no exact solution was obtained in the 3D case. All the results listed in Table 2 were obtained by various approximate methods that, as a rule,

Table 2

Percolation Threshold for 3D Lattices

Lattice type	x_b	x_s
Simple cubic	0.25	0.31
Body-centered cubic	0.18	0.25
Face-centered cubic	0.12	0.20
Diamond	0.39	0.43

employ computers. Naturally, certain slight discrepancies are found between the results published in the scientific literature. Table 2 is a selection of results that are the most reliable, from our point of view.

Now the problem is to try and understand, by looking at this table and Table 1 (see p. 116), which summarizes the results for plane lattices, why some lattices have relatively high percolation thresholds, and why other lattices have relatively low thresholds. We begin with the bond problem.

Factors Determining Percolation Threshold in the Bond Problem

If all bonds are intact, each site is connected with z other sites, where the number z of nearest neighbors varies considerably among lattice types. At a fixed fraction x of intact bonds, each site is connected, on the average, with zx other sites. Let us try and test the following hypothesis: Is it realistic for the quantity zx that gives the average number of sites with which each site is connected to carry sufficient information for deciding whether percolation is present or absent in the lattice? Can it be that we need no other information on the properties of a lattice but its coordination number z, and that percolation sets in in all lattices at the same value of zx? It is rather clear that this hypothesis cannot be very accurate. But is it approximately valid?

This is very easy to prove or disprove. Let us find the product zx_b for all lattices with known percolation thresholds of the bond problem. If this product is universal, that is, identical or at least nearly identical for all lattices, the proposed hypothesis holds or holds approximately.

Table 3

The Product zx_b for Different Lattices

Lattice type	z	x_b	zx_b
Plane lattices			
Square	4	0.5	2.0
Triangular	6	0.35	2.1
Honeycomb	3	0.65	2.0
3D lattices			
Simple cubic	6	0.25	1.5
Body-centered cubic	8	0.18	1.4
Face-centered cubic	12	0.12	1.4
Diamond	4	0.39	1.6

The required data are summarized in Table 3. You see that in plane lattices we have

$$zx_b = 2 \tag{1}$$

with an error less than 10%, and in 3D lattices

$$zx_b = 1.5 \tag{2}$$

The hypothesis of the universal mean number of bonds per site required for the onset of percolation is thus not accurate, but it does hold approximately. If we remember that each of the quantities z and x_b varies by at least a factor of two both in the group of plane lattices and in that of 3D lattices, the accuracy with which the product zx_b is constant within each group must be considered remarkably high.

We conclude that percolation threshold of the bond problem can be approximately evaluated if we know the number of nearest neighbors, and use formula (1) for plane lattices and formula (2) for 3D lattices. Percolation threshold of the bond problem is most sensitive to the number of nearest neighbors, but is much less sensitive to all other properties of lattices (e.g. to the number of next-to-nearest neighbors, i.e. the sites that are second in their closeness to a given site after its nearest neighbors).

We have thus arrived at a very simple and at the same time a relatively accurate method of evaluating percolation thresholds of the bond problem applicable to arbitrary lattices.

How to Evaluate Percolation Threshold in the Site Problem

Let us analyze now a similar scheme for the site problem. It will be natural to test first the familiar version, that is, check the variation of the quantity zx_s among different lattices. It will be readily found that the product varies almost as each of the quantities z and x_s does independently. One should not be surprised: the product zx_b has a well-defined physical meaning for the bond problem, namely, that of the mean number of intact bonds per site. In the case of the site problem, a bond is effective if it connects two white sites, but is ineffective in all other cases. Consequently, the product zx_s is virtually meaningless.

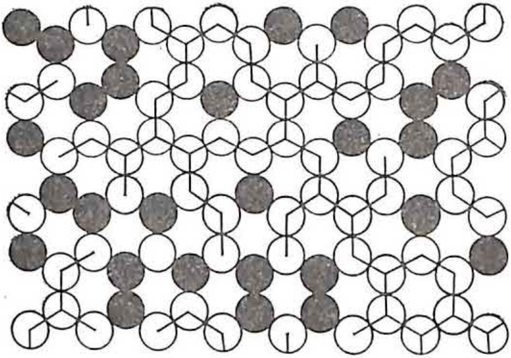


Fig. 31. Construction of tangent circles in a honeycomb lattice. The lattice itself is shown in Fig. 16*c*. The circles have radii equal to one half of the distance to the nearest neighbor. The white sites correspond to white circles, and the black sites to black circles. The percolation paths through white circles are traced with the solid lines.

In 1970 the American physicists H. Scher and R. Zallen proposed a different method for evaluating percolation threshold in the site problem. Their idea was to put each site in correspondence with a specific part of space. Then they went on to say that percolation through white sites appears when the fraction of space occupied by these sites exceeds a certain critical value slightly dependent on the lattice type.

Imagine around each lattice site a sphere (or a circle if we deal with a plane lattice) of radius equal to half the distance to the nearest neighbor. The spheres (circles) constructed around neighboring sites are then tangent to one another (Fig. 31). A white site is assigned a white sphere, and a black site a black sphere. If two white sites are connected, there is a path between them through tangent white spheres (see Fig. 31). Therefore, the onset of percolation means the formation of infinitely long paths through tangent white spheres.

Now we assume that percolation sets in when the fraction of the total volume (or of the total area) occupied by white spheres (or circles in a plane lattice) exceeds a certain critical value independent of lattice type. In order to test this hypothesis, we need to calculate the fraction of volume occupied by white spheres for $x = x_s$ for different lattices with known values of x_s and then compare the results.

First we need to calculate the fraction of volume filled up by white spheres for $x = 1$, that is, in the case when all spheres are white. This quantity is denoted by letter f and referred to as *filling factor*. The filling factor equals the fraction of volume occupied by the spheres constructed around each lattice site and having a radius equal to half the distance to the nearest neighbor. Filling factor essentially depends on lattice type and must be calculated for each specific lattice.

The fraction of volume occupied by white spheres for $x < 1$ is found by multiplying the filling factor by the fraction of white spheres, that is, by x. The fraction of volume filled up by white spheres is thus fx. At percolation threshold it equals fx_s. If the hypothesis on the universality of the fraction of volume at which percolation sets in is correct, the product fx_s must be identical for all lattices.

The filling factors for a number of lattices are listed in the second column of Table 4. In order

Table 4

The Product fx_s for Different Lattices

Lattice type	f	x_s	fx_s
Plane lattices			
Square	0.79	0.59	0.47
Triangular	0.91	0.5	0.46
Honeycomb	0.61	0.7	0.43
3D lattices			
Simple cubic	0.52	0.31	0.16
Body-centered	0.68	0.25	0.17
Face-centered	0.74	0.20	0.15
Diamond	0.34	0.43	0.15

to illustrate how they were calculated, let us determine f for the honeycomb lattice given in Fig. 31. It was shown in Exercise 4 to Chapter 5 that the area per site of the honeycomb lattice is $(3\sqrt{3}/4)\,a^2$, where a is the length of the side of the unit cell. This result has the following meaning: let us draw on a plane, on which the lattice was constructed, a square, a rectangle, a circle, or any other geometric figure, but necessarily such that its size be much greater than the distance between the neighboring sites of the lattice. Divide its area by the number of lattice sites that fell within this figure. The area per site is the limit to which this ratio tends when the figures become infinitely large.

The fraction of area occupied by circles equals the limit of the ratio of the area occupied by the circles to the area of the large figure. The area occupied by the circles equals the product of the number of sites belonging to the large figure by the area of a single circle. In other words, the quantity f equals the ratio of the area of one circle to the area per site.

The circles drawn in Fig. 31 have the radius $a/2$, and hence, their area is $\pi a^2/4$. This gives

$$f = \frac{\pi a^2}{4} \Big/ \frac{3\sqrt{3}a^2}{4} = \frac{\pi}{3\sqrt{3}} \approx 0.605$$

The filling factors for other lattices are calculated in a similar manner; you can see from Table 4 that f varies within a wide range.

The last column of Table 4 lists the products fx_s. You can see that the hypothesis that fx_s is independent of lattice type does not hold too well. However, its variations are small both within the group of plane lattices and within the group of 3D lattices. Hence, to within 10 to 15% the following formulas hold:

$$fx_s = 0.5 \tag{3}$$

for plane lattices, and

$$fx_s = 0.16 \tag{4}$$

for 3D lattices.

The calculation of the filling factor f being relatively simple, formulas (3) and (4) make it possible to evaluate percolation threshold of the site problem for arbitrary lattices.

It is readily understood that the critical fraction of volume occupied by white spheres, at which percolation appears, decreases monotonically with increasing dimensionality of space. In one-dimensional space, that is, in a linear string of sites, percolation through white sites vanishes at no matter how small concentration of black sites. Even a single black site blocks the percolation path because no bypass is possible. The possibility to go around black sites appears in plane (two-dimensional) lattices, and such possibilities become more numerous in 3D lattices because detours are not restricted to a single plane.

The critical volume concept proves fruitful not only in lattice problems. In Chapter 9 we shall encounter a problem in which white and black balls are not located at lattice sites but are randomly poured into a jar. We shall be interested in percolation through white balls in contact with one another. This percolation is also found to arise when the volume occupied by white balls comes to about 0.16 of the total volume. This result changes only slightly if the balls have unequal radii.

Chapter 10 treats the problem of space painted in a random manner by white and black paints. It is found that percolation through regions of one of these colors appears in the 2D case when the fraction of surface area painted white (or black) is exactly equal to 0.5, and in the 3D case when the fraction of volume painted white (or black) is approximately equal to 0.16.

Exercise

1. Check whether the filling factors f listed in Table 4 were correctly calculated.

Chapter 7

Ferromagnetics with Long-Range Interaction. The Sphere Problem

So far, when considering site and bond problems, we have always assumed that each site can be directly connected only with its nearest neighbors, and that the connections between farther removed sites are established through strings of sites each of which is connected with its nearest neighbors. In this chapter the site problem is generalized to the case in which the sites that are not nearest neighbors are directly connected with each other. This problem may be of practical importance, and so it may prove useful to know that it has been sufficiently well elaborated.

If the number of sites with which a given site is connected becomes very large, the site problem transforms to a quite new problem referred to as the *sphere problem*. In fact, this problem plays an especially important role in percolation theory: it is used to understand the transition to metallic conductivity taking place in semiconductors as the impurity concentration in them increases; its solution is at the basis of the theory of hopping conduction in semiconductors, which is an important and interesting phenomenon observed at very low temperatures. Consequently, numerous scientists worked with the sphere problem, and interesting results have been obtained for this and similar problems.

The sphere problem is also interesting because this is the first *nonlattice problem* that we encounter in this book. The random elements with which it operates are defined not at sites of a periodic lattice.

Ferromagnetics with Long-Range Interaction

Let us return to the problem of a ferromagnetic with nonmagnetic impurity atoms, discussed in Chapter 3. There it was assumed that magnetic atoms orient their magnetic moments parallel to each other only if they are nearest neighbors or if they are connected through a string of magnetic atoms that form successive pairs of nearest neighbors. But if all nearest neighbors of a magnetic atom were nonmagnetic atoms, the magnetic moment of this atom was assumed to be arbitrarily oriented.

This model was based on the fact that the interaction between magnetic moments resulting in the parallel arrangement decreases very rapidly with distance; in fact, the decrease is so steep that the magnetic moments that are not nearest neighbors simply "ignore" one another, that is, do not interact at all.

In crystallography, the group of all nearest-neighbor atoms of a given atom is usually called its *first coordination group*, and it has already been mentioned that the number z of nearest neighbors is called the coordination number. The ensemble of equivalent atoms that are next-to-nearest neighbors of the given atom is called the *second coordination group*, and so on. As an example, let us consider the simple cubic lattice (see Fig. 12). The first coordination group in this lattice consists of six atoms located at the edges of the cube projecting from the initial atom. The second coordination group is formed by 12 atoms located at the diagonals of the cube's faces passing through the initial atom. And finally, the third coordination group consists of eight atoms located at the body diagonals of the cube passing through the initial atom.

In Chapter 3 we assumed that the interaction between atomic magnetic moments vanishes beyond the first coordination group. Under this assumption, the calculation of the critical concentration of magnetic atoms at which spontaneous magnetization appeared (or vanished) reduced to the site problem with bonds only between nearest neighbors.

The assumption of short-range interaction does not always hold, so that it is justifiable to discuss a problem in which the interaction between magnetic moments covers several coordination groups, and to determine the critical fraction of magnetic atoms resulting in spontaneous magnetization.

This problem reduces to the site problem in which bonds are formed not only between nearest neighbors. In fact, its formulation contains nothing new. Sites may be white and black (magnetic and nonmagnetic). Two white sites are regarded as connected if the bonds connect the coordination groups in which they are located. If site A is connected with site B, and site B with site C, then A is connected with C. The ensemble of connected sites forms a cluster. The percolation threshold is defined as the fraction of white sites at which an infinite cluster is born.

When bonds are made to extend to farther coordination groups, the percolation threshold x_s must

Table 5

Percolation Threshold in the Site Problem with Bonded Nonnearest Neighbors*

Lattice type	Z	x_s	Zx_s
Plane lattices			
Honeycomb, 1	3	0.700	2.10
Square, 1	4	0.590	2.36
Triangular, 1	6	0.500	3.00
Square, 1, 2	8	0.410	3.28
Triangular, 1, 2	12	0.295	3.54
Honeycomb, 1, 2, 3	12	0.300	3.60
Square, 1, 2, 3	12	0.292	3.50
Triangular, 1, 2, 3	18	0.225	4.05
3D lattices			
Diamond	4	0.425	1.70
Simple cubic, 1	6	0.307	1.84
Body-centered, 1	8	0.243	1.94
Face-centered, 1	12	0.195	2.34
Body-centered, 1, 2	14	0.175	2.45
Simple cubic, 1, 2	18	0.137	2.47
Face-centered, 1, 2	18	0.136	2.45
Simple cubic, 1, 2, 3	26	0.097	2.52
Body-centered, 1, 2, 3	26	0.095	2.47
Face-centered, 1, 2, 3	42	0.061	2.56

---

* The listed thresholds for a single coordination group are not always exactly equal to the data given in the preceding tables. The reason for this is that those tables gave the data that we considered the most reliable, while the present table compares the data calculated by the same method.

obviously diminish. The greater the number of bonds originating at a given white site, the greater the probability that at least one of these bonds will lead to another white site, and correspondingly, the smaller the number of white sites necessary to ensure percolation.

Clearly, this problem is not a bit simpler (rather, it will be more complex) than the standard site problem. However, a number of problems of this type was solved by various approximate techniques, and the results of one of them are listed in Table 5. The first column of the table gives the lattice type and the number of coordination groups to which the bonds were extended. The second column shows the number of sites, Z, with which each site is connected, that is, the total number of sites located in the considered coordination groups (in the case of a single coordination group it coincides with the coordination number z).

The last column of the table gives the product Zx_s. This product, as we indicated in the preceding section, is strongly dependent on lattice type in the case of the site problem with bonds only in the first coordination group. However, as we find in Table 5, at high values of Z it changes the slower, the greater Z is. This is especially clear for three-dimensional lattices where large values of Z are used. Obviously, the product Zx_s tends to a number of the order of 2.6-2.7 independent of lattice type.

In the 3D case the number 2.7 is now considered the most likely value (to within ± 0.1) of the quantity B_c defined as the limit of Zx_s for very large Z:

$$B_c = \lim_{Z\to\infty} Zx_s \qquad (1)$$

It will be explained in the subsequent sections why this limit exists and why it is independent of lattice type, being only a function of the dimensionality of space, that is, of whether the lattice under consideration is plane or three-dimensional.

In order to understand this feature, it is necessary to study the so-called sphere problem.

Exercise

1. Find the location of 42 sites belonging to the first three coordination groups of an fcc lattice.

The Sphere (Circle) Problem

Now we shall consider a different problem or, rather, a problem that seems to be different at first glance. Assume that a plane is covered with circles of identical radius R whose centers are spread on the plane quite randomly and, on the average, uniformly. This means that both coordinates of centers of circles are random numbers distributed uniformly in the range from zero to L, where L is a very large length (in comparison with R) characterizing the size of the system under discussion. The important distinctive feature of this problem is that there is no limitation on the extent to which the circles overlap. The mean number of centers of circles per unit area is

N. In other words, N is the concentration of centers of circles.

Two circles are said to be mutually connected if the center of one of them lies within the other; such circles are sometimes said to be inclusive. If circle A is connected with circle B, and B is connected with C, then A is connected with C. Therefore, too far removed circles can be connected via a string of inclusive circles (Fig. 32).

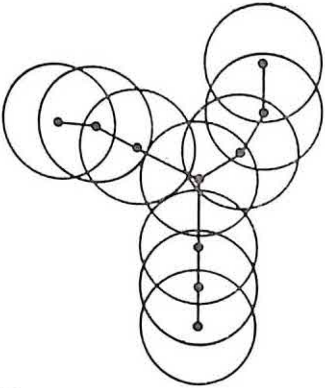


Fig. 32. Percolation paths through inclusive circles are traced by the broken lines. The dots mark the centers of circles.

The problem now is to find the critical concentration N at which percolation sets in through inclusive circles, that is, paths are formed traversing the whole system and composed of inclusive circles. (In other words, an infinite cluster is formed out of mutually connected circles.)

The problem involves two parameters: the concentration N and radius R. (We have left out the size of the system, L, but it is clear that if the system is sufficiently large, the critical value N is almost independent of L.) It is not difficult, however, to ascertain that the presence or absence of percolation depends not on two parameters but only on one, namely, on the dimensionless product NR^2. (The dimensionality of concentration in a plane problem is cm^{-2}.) Let us choose for this parameter the mean number of centers of circles within a circle. It equals

$$B = \pi N R^2$$

The following arguments will immediately show that percolation sets in at a certain value of the parameter B, regardless of the specific values of N and R making up B. Imagine looking at a plane with the circles drawn on it. Let us magnify this picture severalfold, for instance, through a projector. This will be a transformation varying N and R but leaving B invariant because the mean number of centers of circles within a circle will not be changed by this magnification.

It is also easy to understand that this transformation does not affect percolation. If there was no percolation in the initial picture, there will not be any in the magnified picture, and conversely, if there was percolation through inclusive circles in the initial picture, this percolation will not vanish because of magnification.

The transformation changing N and R but not B thus does not affect percolation. Consequently, it depends only on the value of the parameter B whether there will be or not be percolation in a system. Percolation is present when B is large, and absent when it is small.

This new problem that has just been formulated above is referred to as the *circle problem*. Its three-dimensional analogue is called the *sphere problem*. The sphere problem is formulated as follows. The coordinates of centers of spheres of radius R are generated in the three-dimensional space by a suitable random-number generator. Two spheres are said to be connected (or inclusive) if the center of one sphere lies within another sphere. We need to determine the critical concentration of centers at which percolation sets in through inclusive spheres.

It will be readily understood that, as in the plane problem, the presence or absence of percolation is determined exclusively by the value of the parameter B which is the mean number of centers of spheres within a sphere:

$$B = (4/3)\,\pi N R^3$$

where N is the mean number of centers of spheres per unit volume (the dimensionality of volume concentration is cm^{-3}).

We have already mentioned that the sphere problem has important applications to the theory of electric conduction in semiconductors at low temperatures. For this reason, it was analyzed by a number of authors resorting to most various techniques. By today's data, the critical value B_c at which percolation through spheres sets in is 2.7 ± 0.1. The circle problem was not studied as thoroughly, so that a substantial spread is found among the results published by different authors. It appears that $B_c = 4.1 \pm 0.4$.

The Circle (Sphere) Problem Is the Limiting Case of the Site Problem

Let us return to the site problem in which not only nearest neighbors are connected, and explain why the limit on the right-hand side of formula (1) exists, why it is independent of lattice type, and finally, why we denote it by the same symbol B_c as the threshold values of the circle and sphere problems.

We begin with slightly reformulating the site problem. To be specific, we first discuss plane lattices. Generalization to 3D lattices proves to be quite simple.

Let us surround each white site with a circle of radius R that was chosen to be greater than the distance from this site to the sites of the farthest coordination group with which the given site is connected, but smaller than the distance to the sites of the next coordination group. We assume that two circles are connected if the white sites around which they were drawn are connected. This means that two circles are connected if the center of one of them lies within the other, that is, if the circles are inclusive.

The formation of an infinite cluster of connected white sites is equivalent to the onset of percolation through inclusive circles. In other words, percolation through the circles of radius R drawn around white sites sets in at a critical fraction of white sites, x_s (Fig. 33).

The meaning of the quantity Zx is easily understood. The quantity Z is the number of sites (both black and white) that lie within a circle. The product Zx is the mean number of centers of other circles that lie within a circle (or the mean number of white sites within a circle). The quantity Zx_s is the mean number of centers of circles within a circle at which percolation sets in, that is, at which infinite paths can be traced through inclusive circles.

It is now obvious that the product Zx_s has the same meaning as the quantity B_c in the circle

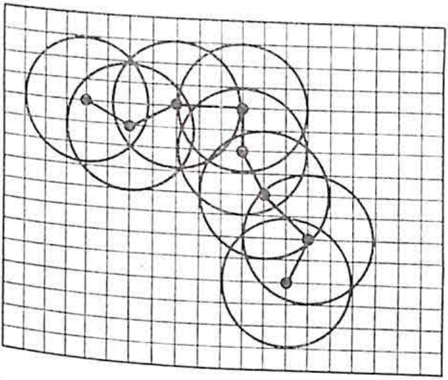


Fig. 33. Percolation path through inclusive circles on a square lattice. The interaction is taken into account at a distance of up to three distances to the nearest neighbors. The percolation path is traced by the broken line.

problem. Quite likely, now it may be more difficult for the reader to point out the difference between the site and circle problems than to notice the similarity. But the difference is there, and it is an important one. The fact is that in the circle problem the points that can serve as the centers of circles can lie anywhere on the plane, while in the site problem these are only the sites of the lattice in question (see Fig. 33). If the number of sites within a circle is small, the difference between the two problems becomes very important. It is then natural that the critical value of Zx_s is a function of lattice type. The total number of sites lying within a circle is Z. As we see from Table 5, Zx_s indeed varies among lattices when Z is not too large.

However, if Z is large, the difference between the two problems fades away. Imagine that we have started with the circle problem, but then shifted the center of each circle to the lattice site nearest to it. This is already a site problem. If the number of lattice sites within a circle is very large, this shift will not, with a high probability, connect the circles that were not connected before the shift, and vice versa. This argument shows that the site and circle problems become equivalent for $Z \to \infty$. The value of B_c determined by formula (1) is independent of lattice type and coincides with the quantity B_c defined in the circle problem.

This chain of arguments can be carried over entirely to the 3D case. The quantity B_c defined by formula (1) coincides for 3D lattices with B_c defined in the sphere problem.

To recapitulate, the site problem on any plane lattice is reducible at large Z to the circle problem, and the site problem on a 3D lattice is reducible to the sphere problem. The limit in formula (1) is, therefore, indeed independent of lattice type but depends on the dimensionality of the space on which the problem has been formulated. (The values of B_c are different for the circle and sphere problems!)

Chapter 8

Electric Conduction of Impurity Semiconductors. The Sphere Problem

Percolation theory proved extremely useful for the understanding of the processes in doped semiconductors (called *extrinsic semiconductors* to distinguish them from pure ones, called *intrinsic semiconductors*). It can rightly be said that at the present time impurity semiconductors represent one of the main fields to which percolation theory is applied. Several chapters of this book deal with impurity semiconductors. The present chapter begins with an exposition of current notions of semiconductor science, briefly interrupting the story of percolation theory.

Intrinsic Semiconductors

Let us start to discuss pure semiconductors, choosing as a basis semiconducting elements of the fourth group of the periodic table, such as germanium and silicon. These elements crystallize into diamond lattice (see Fig. 30*a*, *b*). The four electrons making up the outer shell of each atom form bonds with the four nearest neighbors. The electrons move in such a way that their density is distributed nonuniformly around the atoms, being concentrated in four "strands" stretching from the centers of tetrahedrons, in which the atoms are located, to the vertices of these tetrahedrons. It is these strands that bind up the atoms of the crystal and do not let them fly apart.

It can be imagined, by way of illustration, that each atom has four hands and holds hands with its four neighbors. Each electron is then strongly bound to a neighbor, so that a moderately strong electric field applied to the semiconductor cannot generate an electric current: the number of electrons is exactly equal to that necessary to form the bonds, while the energy required to release a bonding electron is quite high. This energy, or rather, the minimum work that has to be done to transfer an electron from the bound state to the state in which it can move freely through the crystal is called the *energy gap width* (or simply *gap width*) and is denoted by E_g.

The gap width is an extremely important characteristic of a semiconductor, determining to a large extent all its electrical properties. For instance, let us consider electric conduction. The electric conduction in pure (intrinsic) semiconductors exists because a certain fraction of bonds are broken. A bond breakdown releases an electron ready to carry electric current and a "crippled", "three-armed" atom. This "cripple" is also a charge carrier, only its sign is opposite to that of the electron. Indeed, an electric field can transfer an electron from a neighbor to the "three-armed" atom, changing its energy only slightly. As a result, another atom becomes "three-armed". This process will continue, and you easily understand that in contrast to electrons moving in the field, say, from left to right, a "three-armed" atom moves from right to left. (It is important to realize that the atoms themselves do not move. In fact, a moving electron turns the atom which it abandons into a "three-armed" atom.) For this reason, a "three-armed" atom can also be regarded as a charge carrier but, in contrast to the electron, it must be assigned the positive, not negative, charge.

The microscopic picture is somewhat different in other semiconductors consisting of atoms with a different valence, but in all cases the breakdown of a bond produces two charge carriers with opposite signs. One of them (the negative) is called the electron, and the other (the positive) is called the *hole*. A "three-armed" atom is thus a particular case of hole.

The energy required to break down a bond is supplied from the energy of thermal motion of the atoms. As it is proved in statistical physics, at high temperatures the mean energy of thermal motion per atom that performs small-amplitude vibrations around the equilibrium position is $3kT$, where T is temperature in K, and k is the Boltzmann constant ($k = 1.38 \cdot 10^{-16}$ erg/K).

As a rule, the gap width E_g is measured in electron volts (eV). One electron volt is the work done by an electron traversing a potential difference of 1 V: 1 eV $= 1.6 \cdot 10^{-19}$ J $= 1.6 \cdot 10^{-12}$ erg. In germanium $E_g = 0.7$ eV, and in silicon $E_g = 1.1$ eV. At room temperature (300 K) the energy $3kT$ is only 0.08 eV. This is much less than necessary for a bond breakdown both in germanium and silicon.

However, thermal motion is chaotic, and at some random moments the energy of motion of an atom may be very high. This happens extremely rarely to each individual atom, but the total number of atoms is very large. There are roughly

10^{22} atoms in 1 cm^3. (The reader may remember that the number of atoms per mole equals Avogadro's number, $\approx 6 \cdot 10^{23}$.)

For this reason, the concentration of electrons and holes released by broken bonds is much smaller than the concentration of atoms; nevertheless, it is so high that it can result in an appreciable electric conductivity in a certain temperature range. Calculations show that in germanium the concentration of electrons is about $10^{13}\ \text{cm}^{-3}$ at $T = 300$ K. In silicon, where the gap is wider, the concentration at the same temperature is much lower ($10^{10}\ \text{cm}^{-3}$). As temperature is reduced, the concentration of electrons decreases sharply, and the resistance of an intrinsic semiconductor undergoes a corresponding sharp enhancement.

The number of charge carriers at a given temperature is the smaller, the wider the energy gap is. The difference between a semiconductor and a dielectric is merely the width of the gap. As a rule, materials with E_g of about 5 eV or more are referred to as dielectrics. They contain practically no charge carriers at room temperature.

While the difference between dielectrics and semiconductors is more quantitative than qualitative, the difference between metals and dielectrics is a principal one. The energy gap of a metal is zero, and the concentration of charge carriers is high even in the immediate vicinity of the absolute zero of the temperature scale.

Impurity Semiconductors

Let us assume that germanium or silicon is doped with an impurity element from the fifth group of the periodic table, such as phosphorus, antimony, or arsenic. An atom of these elements has five electrons on the outer shell. If such an atom substitutes, for example, a germanium atom located at the center of a tetrahedron, four of its electrons go to form the bonds with the four neighbors and the fifth electron remains dangling. This electron is bound to its atom because if the electron moves away, the atom becomes positively charged and attracts the electron back. Therefore, the electron can be removed to a considerable distance away from the atom only by doing a work against attractive forces. This work is what we call the *energy of bonding* of the electron to an atom.

As we will show below, the bonding energy of this extra electron is relatively low, so that the distance between this electron and the impurity atom is rather large (in comparison with the period of the lattice). As a result, the structure of the impurity atom resembles that of the simplest among atoms, namely, the hydrogen atom. We remind the reader that the hydrogen atom consists of a positively charged heavy nucleus and a light negative electron, the size of the nucleus being negligibly small in comparison with the distance from the nucleus to the electron.

An impurity atom has a similar structure. The unit that acts as the nucleus of a hydrogen atom is not just the nucleus of the impurity atom but includes the inner-shell electrons and the four electrons that form the bonds. We shall be able to show now that the size of this complex is small compared with the distance to the extra electron; the charge of the complex is positive and equals in magnitude the electron charge.

The electron in a hydrogen atom is known to be at a distance of the Bohr radius a_B from the nucleus:

$$a_B = \frac{\hbar^2}{me^2} \approx 0.53 \cdot 10^{-8} \text{ cm} \tag{1}$$

Here $\hbar = 1.05 \cdot 10^{-27}$ erg·s is Planck's constant (divided by 2π), $m = 9.8 \cdot 10^{-28}$ g is the electron mass, and $e = 4.8 \cdot 10^{-10}$ CGSE units is the electron charge.

The fundamental constant $\hbar$ was introduced by the German physicist Max Planck in 1901 in connection with the hypothesis on the quantum nature of electromagnetic waves. This constant enters into the quantum-mechanical equation describing the motion of an electron around the nucleus. The only other constants in this equation are e and m. The Bohr radius a_B is the only possible quantity with the dimensionality of length that can be composed of e, m, and $\hbar$.

The bonding energy of the electron in the hydrogen atom equals

$$E_B = \frac{me^4}{2\hbar^2} \approx 13.6 \text{ eV} \tag{2}$$

The following interpretation of this formula by the order of magnitude is possible. The positively charged nucleus creates a potential e/r at a distance r. At a distance of order a_B it equals e/a_B, and at infinity it is zero. Consequently, the work that has to be done to transfer an electron located at a distance of order a_B from the nucleus to an infinitely large distance equals, by the order of magnitude, e^2/a_B (the potential difference times the electron charge).

Let us return to the impurity atom. The formulas written for the hydrogen atom have to be modified in order to take into account that the extra electron moves not in a vacuum but in a semiconductor crystal. This modifies the form of Coulomb's law. The force acting on an electron separated by a distance r from the nucleus is now $e^2/\varepsilon r^2$ (in CGSE units), where ε is the dielectric constant, or permittivity, of the crystal, and not e^2/r^2. This complication can be taken into account by replacing the quantity e^2 by e^2/ε in the expressions for a_B and E_B. Furthermore, it should be borne in mind that the mass describing the motion of an electron through the crystal does not equal that of a free electron, m.

The point is that the nuclei of the semiconductor's atoms and the inner-shell electrons produce in the crystal a periodic electric potential. One of the most interesting conclusions of the quantum theory of solid state is that the electron very nearly "overlooks" this potential if it is exactly periodic. The qualification "very nearly" consists in that the neglection of the periodic potential in the equations of motion of the electron must be simultaneously accompanied by replacing the electron mass m by a mass m^* which is a function of crystal properties. The quantity m^* is called the *effective mass*.

An impurity atom produces a nonperiodic potential that can by no means be omitted. But, when the motion of the electron in this potential is described, we can neglect the periodic potential of the crystal provided we replace m by m^*.

The equation describing the motion of the extra electron in a crystal around a charged impurity atom thus differs from the equation of motion of an electron in the hydrogen atom by the replacements $e^2 \to e^2/\varepsilon$ and $m \to m^*$.

Let us denote the characteristic distance to which the extra electron is removed from the impurity atom by a_B^*, and the bonding energy of this electron by E_B^*. Making use of formulas (1) and (2), we obtain

$$a_B^* = 0.53 \cdot 10^{-8} \left(\frac{m}{m^*}\right) \varepsilon \text{ [cm]} \tag{3}$$

$$E_B^* = 13.6 \left(\frac{m^*}{m}\right) \frac{1}{\varepsilon^2} \text{ [eV]} \tag{4}$$

As a rule, the effective masses in semiconductors are substantially smaller than the mass of the free electron, and the dielectric constant equals 10-20 (e.g. in germanium $m^* \approx 0.1\ m$ and $\varepsilon = 16$). Consequently, the characteristic distance at which the extra electron is found equals, in typical semiconductors, from several tens to several hundreds of angstroms (1 Å = $= 10^{-8}$ cm), which is much greater than the interatomic distance (e.g. in germanium this spacing is 2.45 Å).

We have thus obtained that the extra electron in an impurity atom is removed to a considerable distance from the atom and is kept at this distance by the attractive forces caused by the nucleus. The fact that the distance a_B^* is large is already a sufficient indication of smallness of the bonding energy E_B^*. Indeed, the data cited above show that in typical semiconductors the bonding energy E_B^* is from several hundredths to several thousandths of an electron volt; hence, it is much smaller than the gap width E_g (e.g. in germanium $E_g = 0.7$ eV).

This is natural because it is easier to break loose the extra electron off the impurity atom than the electron bonding the host atoms. The thermal energy kT reaches the level of 0.01 eV at a temperature of the order of 100 K. As a rule, at this temperature a considerable fraction of extra electrons separate from their impurity atoms and take part in the transfer of electric current. Therefore, impurity atoms belonging to the fifth group of the periodic table let go off their extra electrons rather easily. For this reason, they were given the name *donors*.

Suppose that the impurity atoms belong to an element of the third group of the periodic table, for example, boron, aluminium, gallium, or others. These atoms have three electrons on the outer shell, so that they *lack* one electron for forming the bonds with their four neighbors. This electron is readily borrowed from the neighboring atoms of the semiconductor, but then one of the neighbors becomes "three-armed" or, in other words, a hole appears in the neighborhood of the impurity atom. The impurity atom captures a fourth electron and thereby becomes negatively charged. The hole is thus attracted to this atom by electric forces, and a work has to be done to pull the hole away. This work is called the bonding energy between the hole and the impurity atom.

The calculation of the bonding energy of the hole again leads to the problem of a hydrogen-like atom, but in this last case it is a positively charged hole that moves around a negatively charged nucleus. The role of the nucleus is now played by an impurity atom together with the captured additional electron. This additional electron turns into a strand forming an interatomic bond, so that the region in which it moves does not exceed one interatomic distance. At the same time, a hole is bonded much less strongly to an impurity atom. A distance from the atom to the hole, and the bonding energy, are given by formulas (3) and (4). We only have to take into account that the effective mass of the hole that must be used in this case differs, generally speaking, from the electron effective mass. As a rule, this mass is also considerably smaller than the free electron mass, so that a hydrogen-like atom with a hole also has a size of the order of tens of angstroms, and its bonding energy is of the order of several hundredths of one electron volt.

At temperatures of about 100 K the thermal motion breaks the bonds between holes and impurity atoms, after which the holes begin an "independent way of life" and, when an electric field is applied, participate in the electric current.

Impurity atoms from the third group of the periodic table thus readily accept an electron and form a hole; hence, the term *acceptors* for such dopants.

Let us summarize this section. Mobile charge carriers are formed in semiconductors only at the expense of the thermal motion energy. They can be formed when bonds of the lattice are broken. This requires that a work be done equal to the energy gap width E_g. The outcome is the simultaneous appearance of an electron and a hole. An electron or a hole can also be born individually when their bond to an impurity atom is broken. The bonding energy between an electron or a hole and an impurity atom being much smaller than the gap width, the probability for the electron to break away from the impurity atom is much higher than the probability of breaking a lattice bond. On the other hand, the concentration of impurity atoms is usually lower by many orders of magnitude than the concentration of lattice atoms. Consequently, when temperature is increased, first of all charge carriers on impurity atoms are separated, the concentration of intrinsic carriers (electrons and holes born of broken lattice bonds) being negligibly small. This temperature range is called the range of extrinsic conduction. However, as the temperature is further increased, the concentration of intrinsic charge carriers becomes comparable with that of impurity atoms and then grows larger than it. This range is called the range of intrinsic conduction.

The conclusion which is the most important for further exposition is that at a very low temperature at which the thermal motion energy kT is small in comparison with the bonding energy of electrons to impurity atoms, the *semiconductor has neither extrinsic nor intrinsic charge carriers.* The bonds are intact where they ought to be, and impurity electrons and holes are localized in the vicinity of their atoms. In this range, electric conduction of the semiconductor drops very sharply to zero as its temperature is lowered. This behavior distinguishes a semiconductor from a metal in which the concentration of mobile charge carriers remains high no matter how small the temperature is.

Transition to Metallic Electric Conduction at Increased Impurity Concentrations

This principal difference between semiconductors and metals suddenly disappears when impurity concentration is increased. If concentration goes beyond a certain critical value N_c, electric conductivity remains relatively high and weakly dependent on temperature no matter how low the temperature is. This electric conduction is referred to as *metallic*. This does not mean that the electric conductivity of a semiconductor becomes comparable with that of good metals. Far from it: the electric conductivity of a semiconductor always stays many orders of magnitude lower. The term only reflects the behavior of electric conductivity at low temperatures. *The transition to metallic electric conduction that takes place at an increased impurity concentration is called the metal-insulator transition*, or *the Mott transition* (after the famous British physicist Sir Nevill Mott).

Experiments have demonstrated that the critical impurity concentration N_c at which the Mott transition takes place varies quite considerably among different semiconductors. A reliable estimate of the quantity N_c is given by a relation

$$N_c a_B^{*3} \approx 0.02 \tag{5}$$

For instance, in germanium $N_c = 10^{17}$, and in silicon $N_c = 3 \cdot 10^{18}$ cm^{-3}.

Even now we lack a satisfactory mathematical theory of the Mott transition in semiconductors. In fact, this is one of the most complicated problems in solid state theory. The qualitative picture looks as follows.

The structure of impurity atoms resembles that of univalent elements of the first group of the periodic table (H, Li, Na, K). Like these elements, an impurity atom has a single electron on its outer shell. When crystallized, elements of the first group form good metals. Hydrogen forms a molecular crystal which is an insulator. However, there is every reason to believe that this insulator also converts to a metal at a sufficiently high pressure. (The research in this field is conducted nowadays with such vigor that it could make the topic for a separate book.)

It appears as nearly obvious that if hydrogen-like impurity atoms are distributed in a semiconductor with a sufficiently high density, they will also form a metallic system.

It might seem strange at first glance that this argument has a bearing on the Mott transition. Indeed, the concentration of atoms in a metallic sodium is of the order of 10^{22} cm^{-3}, that is, greater than the concentration of impurities in germanium, at which the Mott transition occurs, by a factor of about 100 000. Obviously, sodium cannot be a metal at such low concentrations.

However, it is not just the concentration but the degree to which the electron shells of neighboring atoms overlap that determines whether a system is metallic or not. If the atoms are at such a large distance that the regions of space in which their outer-shell electrons move are far from one another, these are simply isolated atoms. But if the nuclei of the atoms are sufficiently close for these regions to overlap, the atoms lose their individuality. Their outer-shell electrons "cannot know" to which of the nuclei they belong. A system of a large number of atoms appropriates, or collectivizes, electrons. These electrons form a separate system capable of conducting an electric current. This material is a metal.

The degree of overlapping is determined for hydrogen-like atoms by a dimensionless parameter Na_{B}^{*3}. (The quantity $(4/3)\,\pi N a_{\mathrm{B}}^{*3}$ is the mean number of atomic nuclei within a sphere of radius equal to the effective Bohr radius.) It has already been mentioned that since an impurity atom is located inside a semiconductor, its radius is anomalously large. For this reason, condition (5) is met already at $N = 10^{17}\ \mathrm{cm}^{-3}$ (for germanium).

It must be clear by now from the arguments given above why the Mott transition occurs at approximately the same value of the parameter Na_{B}^{*3} in all conductors, although critical concentrations N_{c} may vary quite considerably. Indeed, it is this parameter that determines the overlapping of neighboring atoms!

The following question can be posed: What should be the concentration of atoms in a hydrogen crystal for the overlapping of its atoms to be the same as the overlapping of impurity atoms in a semiconductor at the concentration corresponding to the Mott transition? The overlapping is given by formula (5), but now it is natural to substitute into it the Bohr radius found from formula (1). This will give an enormously high concentration:

10^{23} cm^{-3}. Hence, the overlapping corresponding to the Mott transition is extremely large.

There is no ground, therefore, to be surprised that at a concentration greater than N_c impurity atoms form a metallic system. This fact lies at the foundation of the modern theory of semiconductors with a high impurity concentration.

The Mott Transition and Sphere Problem

Percolation theory offers a simplified description of the Mott transition which takes into account that impurities are located in a semiconductor randomly and can form tighter or looser ensembles. Assume that impurity atoms in a certain region are so close to one another that their electron shells strongly overlap and the outer-shell electrons become "collective property". This region is a piece of metal: a potential difference applied to this region would cause an electric current.

The presence of such regions is not sufficient, however, for a large sample to behave as a metal. If metallic regions are infrequent, they are not in contact, forming only isolated metallic islands in the material that at low temperatures behaves as a dielectric. Taken as a whole, the combination is equally dielectric.

As impurity concentration increases, the fraction of space occupied by metallic regions grows, and at a certain critical concentration N_c the metallic regions form a connected system of "lakes and canals" that permeates the whole semiconductor crystal. Beginning with this concentration, the electric conduction of a large sample becomes metallic.

Obviously, the concepts presented above must be formulated mathematically in terms of percolation theory, although the formulation will not come easily. The main difficulty is that we do not know the impurity concentration at which a region can be regarded as metallic.

The simplest model of the Mott transition enunciated in the sixties reads as follows. Imagine that each impurity atom is a metallic ball

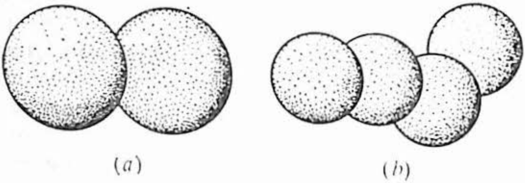

(a) (b)

Fig. 34. (a) Overlapping spheres representing atoms with collectivized electrons; (b) a string of overlapping spheres forming a metallic channel through which electric current can flow.

of radius r_0. The balls can penetrate into one another (Fig. 34*a*), that is, the regions of space that they occupy can overlap. The balls form strings (Fig. 34*b*) and regions of more complex shape that are regarded as metallic by definition. We want to find the concentration N_c of balls beginning with which the metal regions make the whole large sample electrically conductive.

But how to choose the radius r_0? A hydrogen-like impurity atom does not have well-defined boundaries. The probability of finding an electron at a distance r from the nucleus diminishes with increasing r by the law $\exp(-2r/a_B^*)$, where $e = 2.718$ is the base of natural logarithm. The probability of finding an electron at $r = a_B^*/2$ is less by a factor of 2.7 than that at $r = 0$; at $r = a_B^*$ it is less by a factor of 7.4 than that at $r = 0$; at $r = 1.5\, a_B^*$ it is smaller by a factor of 20. The effective radius of the atom must definitely be proportional to the length a_B^*:

$$r_0 = q a_B^* \tag{6}$$

This is an important statement. It signifies that the numerical coefficient q must be fairly universal: it varies rather weakly in going from one semiconductor to another, while the length a_B^* varies quite considerably.

It is very difficult to calculate q on the basis of physical arguments. A simpler approach is to find the critical concentration N_c corresponding to the Mott transition in terms of the model formulated above. This will give us N_c as a function of the ball of radius r_0. Then, making use of formula (6), we must express N_c in terms of q and a_B^*. According to experimental data, N_c is given by formula (5). Comparing the theoretical expression with formula (5), we can find q, that is, determine it "as if" from experimental data.

Let us start the realization of this program. We have to solve the following problem of percolation theory. Spheres of radius r_0, whose centers are distributed in space randomly and on the average uniformly, are drawn in the three-dimensional space. The mean number of centers of spheres per unit volume is N. Two spheres are regarded as connected if they overlap (see Fig. 34b). We need to find the critical concentration N_c at which percolation sets in through overlapping spheres, that is, paths appear that traverse the whole system and are composed of overlapping spheres (see Fig. 34*b*).

This problem differs from the sphere problem formulated in the preceding chapter in that the connected spheres in the sphere problem were not merely overlapping but inclusive, that is, their centers were at a distance smaller than r_0, not than $2r_0$. However, this difference is not too significant, and the results of one problem are easily transferred to the other. Indeed, if an infinite path exists through inclusive spheres of radius $2r_0$ at a certain concentration of centers of spheres, there also exists a path through overlapping spheres of radius r_0. This is just the same path, that is, the path through the same

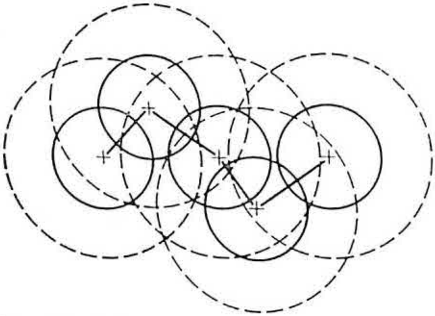


Fig. 35. String of inclusive circles of radius $2r_0$ is shown by the dashed lines. The centers of circles are marked by the crosses. The percolation path through these centers is shown by the broken line. The percolation path through overlapping circles of radius r_0 follows the same centers.

centers. This argument is illustrated in Fig. 35 for the plane problem. (The reader will easily visualize the corresponding drawing for the 3D case.) If there is *no* path through inclusive spheres of radius $2r_0$, there is *no* path through overlapping spheres of radius r_0. Hence, the critical concentration corresponding to the percolation threshold through inclusive sphere of radius $2r_0$ equals the critical concentration corresponding to the percolation threshold through overlapping spheres of radius r_0.

According to the results given in the preceding chapter, the critical concentration for percolation through inclusive spheres of radius $2r_0$ is determined by the condition

$$(4/3)\,\pi N_{\rm c}\,(2r_0)^3 = B_{\rm c} \approx 2.7 \qquad (7)$$

Substituting (6) into (7), we obtain

$$N_{\rm c} a_{\rm B}^{*3} = \frac{0.08}{q^3} \qquad (8)$$

The coefficient q can now be found by comparing (8) with expression (5) obtained from experimental data. In order for these expressions to coincide, we have to set $q = 1.6$. Therefore, the effective radius r_0 equals $1.6a_{\rm B}^*$. The probability to find an electron at the point at a distance of $r = r_0$ from the nucleus is less by a factor of 24 than the probability to find it at the point $r = 0$.

The most important feature of this distribution of the metal-insulator transition lies not so much in calculating the quantity q that determines the effective atomic radius but, of course, in gaining an insight, through percolation theory, into the internal organization of semiconductors at impurity concentrations close to N_c.

If the suggested description is correct, a semiconductor contains a system of metallic channels that run across the matrix. Electric current flows through these channels as through wires. In Part III of this book it will be shown that the fraction of volume of a semiconductor occupied by these channels is very small if impurity concentration is close to N_c. This entails specific properties of the electric conduction and other important characteristics of semiconductors.

Exercise

1. Indium antimonide (InSb) is a semiconductor with a very narrow gap (0.18 eV at room temperature). Effective masses are also small in such semiconductors. Assuming that the effective electron mass m^* is 0.015 m and the dielectric constant is $\varepsilon = 18$, calculate the effective Bohr radius a_B^* and the critical concentration N_c corresponding to the Mott metal-insulator transition.

Chapter 9

Various Generalizations of the Sphere Problem

Inclusive Figures of Arbitrary Shape

We have already mentioned that the motion of electrons through a crystal is described by the effective mass m^*; this m^* may considerably differ from the free electron mass, that is, the mass of an electron in an empty space. It was also found that the effective mass may differ from the free electron mass not only in magnitude. The point is that not all directions in a crystal are equivalent. For instance, the motion of an electron along the edges of a cube may differ from the motion along body diagonals and from the motion along face diagonals of the cube. Consequently, the effective mass is not necessarily identical in different directions. As a result, the region in which an electron moves in the neighborhood of a donor atom will not be spherical. It may be ellipsoidal or even more complex.

These arguments explain why the sphere problem was generalized to the case of figures of arbitrary shape. The new problem is formulated as follows: the sites (centers) are distributed in space randomly and on the average uniformly. The site concentration is N. The sites are surrounded by identical closed surfaces of arbitrary shape.

The surfaces surrounding different sites are identical not only in shape but also in surface orientation in space. If, for instance, the surfaces are fish-shaped, the tails of all the fishes must face in the same direction.

The volume within one surface is V. Two sites are said to be connected if one of them lies within the surface drawn around the other site (inclusive surfaces). We have to find the critical concentration N_c at which percolation through connected sites sets in.

As in the sphere problem, the presence or absence of percolation is determined only by the value of the parameter B which equals the product VN. This is the mean number of sites within the volume bounded by one surface. This number will not change if we enlarge or contract the scales along all the three directions, that is, if we multiply or divide the coordinates of all sites in the system and of all points on the surfaces by the same factor. It is also clear that this rescaling will not suppress percolation if it existed; neither will it bring it to existence. Therefore, percolation "does not respond" to those changes in N and V that leave B unaltered. Consequently, it will be more convenient, as in the sphere problem, to argue not in terms of the critical concentration N_c but in terms of the critical parameter B_c:

$$B_c = N_c V \tag{1}$$

If we keep the shape of the surface drawn around each site constant but increase the volume V bounded by this surface, say, twice, the critical concentration N_c will be half as large, while the parameter B_c will not be effected. It depends only on the surface *shape*.

Formula (1) generalizes the formula

$$B_c = (4/3)\,\pi N_c R^3 \tag{2}$$

that we used earlier for the sphere problem.

In the general case, the critical value of B_c at which percolation sets in must not be equal to the value of $B_c = 2.7$ obtained for the sphere problem. At present the value of B_c for different figures is a subject of intensive study.

The Ellipsoid Problem

The Soviet scientists Ya. G. Sinai and B. I. Shklovskii proved that certain distinct surfaces have the same value of B_c. Thus, the ellipsoid and sphere (and the ellipse and circle in the plane case) are such surfaces.

First we remind the reader of the definitions of ellipse and ellipsoid. The ellipse is a closed curve drawn on a plane obtained from a circle by extending (or contracting) it along one of its axes (Fig. 36). In order to realize this extension or contraction, it is necessary to transform each point M of the circle with x-, y-coordinates into the point M' with coordinates $x' = kx$, $y' = y$, where k is the extension factor ($k > 1$ corresponds to extension, and $k < 1$ to contraction).

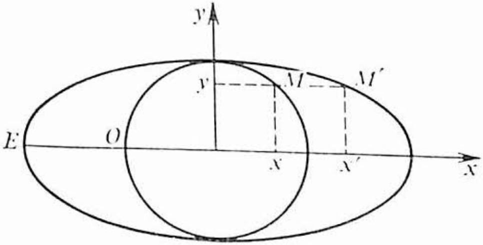


Fig. 36. Ellipse E obtained from circle O by extending it along the x-axis.

The surface obtained from a sphere by extension (or contraction) along one of its axes is called the *ellipsoid of revolution*. In order to realize this extension or contraction along the z-axis it is necessary to transform each point on the sphere with x-, y-, z-coordinates into the point with coordinates $x' = x$, $y' = y$, $z' = kz$. The figure thus obtained is shown in Fig. 37. It is called the ellipsoid of revolution because a rotation by any angle around the z-axis (called the

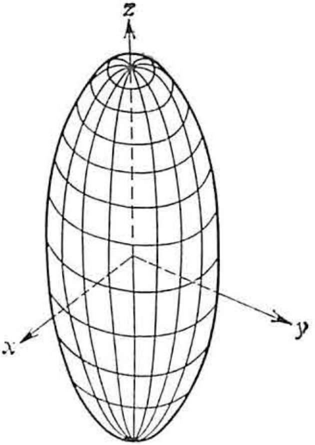


Fig. 37. Ellipsoid of revolution.

axis of revolution) transforms this solid into itself. If it is cut by a plane drawn through the z-axis the section is an ellipse. The section by a plane drawn perpendicularly to the z-axis is a circle.

The general-type ellipsoid is obtained from an ellipsoid of revolution by extending (or contracting) it along one of its axes perpendicular to the axis of revolution. The new extension factor is not necessarily equal to k. The section of the general-type ellipsoid by any plane parallel to the xOy-, xOz-, yOz-planes is an ellipse.

Now we shall prove that ellipsoids and spheres have identical B_c. Let the sites be randomly distributed in space with the mean concentration N. Draw a sphere of radius R around each site. If $B = (4/3)\,\pi N R^3 > B_c^{\mathrm{sph}}$ there is percolation, otherwise, there is no percolation through contacting spheres. Here B_c^{sph} is the critical value of B for the sphere problem ($B_c^{\mathrm{sph}} = 2.7 \pm 0.1$).

Let us apply the extension $y' = y$, $x' = k_1 x$, $z' = k_2 z$ along the z- and x-axes in order to transform the coordinates of both the sites and points on the spheres. (If we dealt with a plane problem, the picture could be modelled by marking the sites and circles on a rubber band which is then extended in one direction. Likewise, we could imagine a three-dimensional "rubber space" in which we have marked the sites and spherical surfaces. Then this "rubber space" is extended in two directions.)

If the x-coordinates of sites are random numbers uniformly distributed within the interval from 0 to L, where L is the size of the system, the new x'-coordinates obtained by multiplying x by k_1 are also random numbers uniformly distributed within the range from 0 to $k_1 L$. The z'-coordinates are uniformly distributed within the interval from 0 to $k_2 L$, while the y-coordinates remain unaltered. The concentration N' of sites is not equal to N.

All spheres now transformed into ellipsoids with volume V'. (It can be shown that $V' = k_1 k_2 \cdot (4/3)\,\pi R^3$. This is irrelevant for the derivation to follow.) As a result, the sphere problem converted to the ellipsoid problem. Here with coordinates $x' = x$, $y' = y$, $z' = kz$. The figure thus obtained is shown in Fig. 37. It is called the ellipsoid of revolution because a rotation by any angle around the z-axis (called the

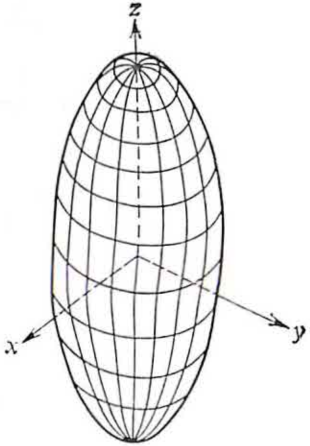


Fig. 37. Ellipsoid of revolution.

axis of revolution) transforms this solid into itself. If it is cut by a plane drawn through the z-axis the section is an ellipse. The section by a plane drawn perpendicularly to the z-axis is a circle.

The general-type ellipsoid is obtained from an ellipsoid of revolution by extending (or contracting) it along one of its axes perpendicular to the axis of revolution. The new extension factor is not necessarily equal to k. The section of the general-type ellipsoid by any plane parallel to the xOy-, xOz-, yOz-planes is an ellipse.

Now we shall prove that ellipsoids and spheres have identical B_c. Let the sites be randomly distributed in space with the mean concentration N. Draw a sphere of radius R around each site. If $B = (4/3)\,\pi N R^3 > B_c^{\mathrm{sph}}$ there is percolation, otherwise, there is no percolation through contacting spheres. Here B_c^{sph} is the critical value of B for the sphere problem ($B_c^{\mathrm{sph}} = 2.7 \pm 0.1$).

Let us apply the extension $y' = y$, $x' = k_1 x$, $z' = k_2 z$ along the z- and x-axes in order to transform the coordinates of both the sites and points on the spheres. (If we dealt with a plane problem, the picture could be modelled by marking the sites and circles on a rubber band which is then extended in one direction. Likewise, we could imagine a three-dimensional "rubber space" in which we have marked the sites and spherical surfaces. Then this "rubber space" is extended in two directions.)

If the x-coordinates of sites are random numbers uniformly distributed within the interval from 0 to L, where L is the size of the system, the new x'-coordinates obtained by multiplying x by k_1 are also random numbers uniformly distributed within the range from 0 to $k_1 L$. The z'-coordinates are uniformly distributed within the interval from 0 to $k_2 L$, while the y-coordinates remain unaltered. The concentration N' of sites is not equal to N.

All spheres now transformed into ellipsoids with volume V'. (It can be shown that $V' = k_1 k_2 \cdot (4/3)\,\pi R^3$. This is irrelevant for the derivation to follow.) As a result, the sphere problem converted to the ellipsoid problem. Here we can introduce a quantity $B' = N'V'$ and find the critical value of B' at which percolation through ellipsoids sets in. Let us denote it by B_c^{ell}.

The rest of the proof separates into the following steps:

1. $B = B'$. All sites that lay within a certain sphere prior to transformation are found inside the ellipsoid obtained from this sphere after the transformation. Indeed, when the sphere is extended (or contracted), the points internal to the sphere remain internal, and those external to it remain external at each step of extension. Hence, the mean number of sites, B, which fell inside one sphere prior to transformation equals the mean number of sites, B', which fell inside one ellipsoid after the transformation.

2. If $B > B_c^{sph}$, then $B > B_c^{ell}$ (3)

If $B < B_c^{sph}$, then $B < B_c^{ell}$ (4)

Indeed, if two spheres were connected prior to transformation, that is, the center of one of them was inside another, the two ellipsoids formed out of these spheres are also connected because the transformation will leave the internal points within, and the external points without, these surfaces. If two spheres were not connected, the ellipsoids obtained from them are not connected either. This yields that if $B > B_c^{sph}$, that is, if infinite percolation paths exist through connected spheres, infinite percolation paths also exist through connected ellipsoids, and this signifies, in turn, that B' is greater than the threshold value B_c^{ell}. This gives us condition (3) because $B' = B$. If $B < B_c^{sph}$, that is, no percolation exists through connected spheres, there is no percolation through connected ellipsoids, that is, $B' < B_c^{\text{ell}}$. This gives us condition (4).

3. Since conditions (3) and (4) must hold for any value of B, they imply that $B_c^{\text{sph}} = B_c^{\text{ell}}$, which was to be proved.

Other Surfaces

The ellipsoid problem is very important in semiconductor physics, but in some cases the ellipsoid is too simple a surface. Physicists were therefore much interested in how the critical value of B_c depends on the shape of the surface when exact relations cannot be obtained. To gain this knowledge, Monte Carlo computer simulations were carried out, and the following surprising and extremely important fact was found out: B_c is only very weakly dependent on the surface shape.

The shapes that were analyzed were a cube and a tetrahedron. It was found that the critical values for these solids do not differ, within the error of calculations (± 0.1), either from each other or from the value of B_c for the sphere problem. Attempts to find a solid with least resemblance to a sphere led to a "three-dimensional cross", that is, a figure formed by three elongated parallelepipeds intersecting at the origin of coordinates (Fig. 38). It was found that B_c for this surface is only 20% less than that for the sphere.

The same conclusion was obtained for plane figures. An analysis demonstrated that B_c for squares differs from B_c for circles by mere 2%.

Consequently, B_c is, to a satisfactory accuracy, universal within a class of figures of identical dimensionality (i.e. either plane or three-dimensional). The reason for this observation may be a certain relation similar to that proved in the preceding section, so that exact equalities in fact hold where computer simulation discovers only a slight deviation or no deviation at all,

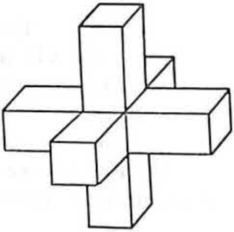


Fig. 38. "Three-dimensional cross".

within computational errors. (Small deviations found by computer calculations must be looked at critically because it is not always easy to carry out a correct evaluation of results.) Unfortunately, this is all we know at present.

Another Experiment at the House Kitchen. The Hard-Sphere Problem

In 1974 three students of the Harvard University in the USA, J.B. Fitzpatrick, R.B. Malt, and F. Spaepen, conducted the following simple experiment. Five thousand small balls, part of them made of aluminium and another part of plastic, were placed in a jar. Before starting the measurements, the balls were thoroughly stirred and then the jar was intensely shaken to make the packing of balls maximally tight. A metal foil electrode was placed at the bottom of the jar, and another electrode was placed on top of the balls, after which the system was compressed by a load of 15 kg.

Aluminium is a highly conductive metal, and the plastic was an insulator. The aim of the experiment was to determine the critical fraction x_c of aluminium balls at which electric current flows between the electrodes, that is, paths appear through the aluminium balls in contact. It was found that $x_c \approx 0.25$.

In addition, the experiment made it possible to study the electric conductivity of the system as a function of x at $x > x_c$.

Here we encounter another problem of percolation theory, namely, the *hard-sphere problem*. The closest relation to this new problem is perhaps the site problem. Let us recall our approach to an approximate evaluation of percolation threshold in the site problem used in Chapter 6. A sphere is constructed around each lattice site, its radius being equal to half the distance to the nearest neighbor. The spheres constructed around white sites are said to be white, and those around black sites are said to be black. Percolation through white sites is equivalent to the existence of percolation paths through tangent white spheres (see Fig. 31).

The difference between the site problem and the new problem lies in that the centers of spheres in the site problem are the sites of a regular lattice, while in the new problem the centers can be located at any point. It will be shown below that this difference is not very important.

The difference between the new problem and the sphere problem discussed in the preceding chapters and the preceding sections of this chapter lies in that the spheres that we operate with in the new problem are assumed to be hard and thus cannot overlap. This is an essential distinction.

It was shown in Chapter 6 that in the site problem percolation through white sites sets in when the fraction of space filled by the spheres constructed around these sites is approximately 0.16. It was found that this number is almost independent of lattice type. It is then natural to conjecture that if it is nearly independent of lattice type, it should not strongly depend on whether *the lattice is or is not there*. If this hypothesis is correct, the fraction of volume filled with metallic balls at which percolation appears through these balls must be approximately 0.16.

Let us denote by f, as in Chapter 6, the filling factor, that is, the fraction of volume occupied jointly by both aluminium and plastic balls. By definition, the quantity x is the ratio of the number of aluminium balls to the total number of balls. Hence, the fraction of volume occupied by aluminium balls equals fx. If this fraction of volume equals 0.16 at percolation threshold, the critical value x_c can be found from the condition $fx_c = 0.16$.

The filling factor of a system of tightly packed but randomly distributed balls is well known. Such a system is familiar to mankind from the times of antiquity. If it was necessary to measure a certain amount of grain or some other free-flowing material, it was poured into a special vessel (a *measure*), shaken, and compacted. In modern science this system is a model of atomic arrangement in *amorphous metals*. Amorphous metals, also referred to as metallic glasses, are materials with metallic electric conductance but having no crystal structure. It was found that the atomic arrangement of amorphous metals strongly resembles the arrangement of tightly packed incompressible balls. For this reason, the properties of tightly packed randomly distributed balls were studied very thoroughly (mostly in computer models). It was found that the fraction of volume filled with the balls is $f = 0.637$.

Let us return to the problem of percolation through metallic balls. Having determined x_c by means of the formula $x_c = 0.16/f$, we obtain $x_c = 0.25$, in complete agreement with the result of the three students.

Experiments on determining the percolation threshold were often repeated, each time with better instruments. Experiments were conducted in which balls in the mixture were of *different* radii. The radii of both metallic and dielectric balls were varied in the mixture in a wide range. It was found that in this case the critical volume fraction fx_c is approximately 0.17, being almost the same, to within experimental accuracy, as in the case of identical balls.

Percolation threshold in the hard-sphere problem can thus be evaluated rather easily after we assume that the critical volume fraction filled with metallic balls is approximately 0.16. Note that this fraction is much greater in the problem of overlapping spheres.

The hard-sphere problem proved very important for applications as well. It accounts for the theory of heterogeneous materials manufactured from an insulator with tiny metallic inclusions. Such materials are now an important object of study. They are prepared and utilized both as thin films and as bulk samples. The electrical properties of such materials in the vicinity of percolation threshold are fascinating. For instance, the capacitance of a capacitor filled with such a material grows to infinity when the fraction of volume filled with the metal tends to percolation threshold. This phenomenon stems from an enormously high mutual capacitance of large metallic clusters. The description of electrical properties of heterogeneous materials is growing nowadays into an independent domain of percolation theory.

Chapter 10

Percolation Level

"The Flood"

Six days and [six] nights
Blows the flood wind, as the south-storm
 sweeps the land.
When the seventh day arrived,
The flood (-carrying) south-storm
 subsided in the battle,
Which it had fought like an army.
The sea grew quiet, the tempest
 was still, the flood ceased.
I looked at the weather: stillness
 had set in,
And all of mankind had returned to clay.
The landscape was as level
 as a flat roof.
I opened a hatch, and light fell upon
 my face.

In: *Ancient Near Eastern Texts Relating to the Old Testament*. Ed. by J. B. Pritchard, Princeton Univ. Press, 1955, p. 94
(*Accadian Myths and Epics, the Epic of Gilgamesh*, Gilgamesh, Tablet XI, translator E. A. Spaiser)

This description of the Flood is found in the Babylonian Epic of Gilgamesh, that most ancient work of literature which dates back roughly to the second millennium BC.

When the waters began to subside, the summits of the highest mountains appeared above the surface. The water subsided lower and lower until it reached its normal level. Just imagine this vast picture: a huge system of mountain ridges, such as the Himalayas, gradually emerged from the water. First the highest peaks broke the surface and formed islands, then the zone of alpine meadows was liberated, and finally, water dropped to foothills.

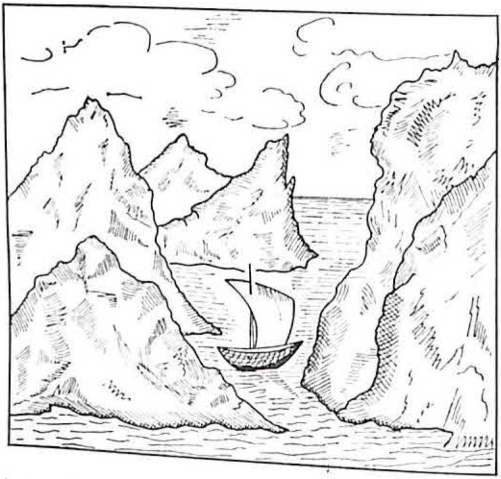


Fig. 39. Voyage during the Flood.

Suppose that we want to answer the question: To what level must the water drop in order for the last waterway across the whole system of ridges to disappear (Fig. 39)? Clearly, such a way exists as long as a certain fraction of moun-

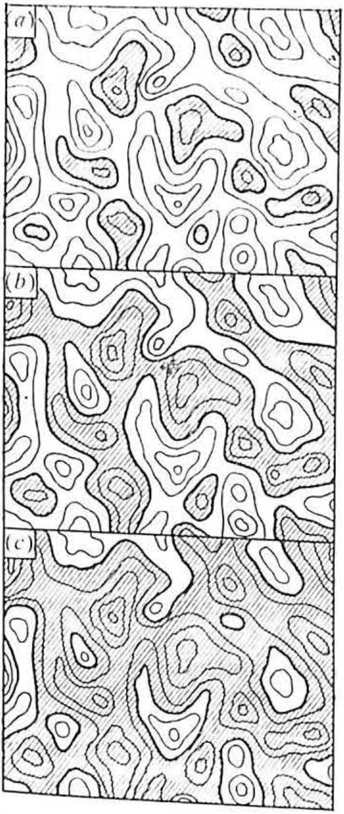


Fig. 40. Map of a mountain system flooded with water. The submerged regions are shaded.

tainous passes remain under water, but then it dries up.

Figure 40 illustrates a geographic map of a mountain system on which the thin lines trace the contour lines (curves of identical height).

The thicker lines mark the contours corresponding to water level. They separate water from dry land. The submerged regions are shaded in the drawing. Figure 40*a* shows the very beginning of the Flood: the lakes do not communicate with one another. Then the water level climbed, and Fig. 40*b* now shows the waterways across the system. And in Fig. 40*c* only individual summits are above the water.

The level of water at which this waterway appears (or disappears) is called *percolation level*. The problem of finding this waterway is a plane problem of percolation theory. It can be reformulated. Let us assume that a plane is randomly painted by white and black paints. Let the fraction of area painted white be x. At small values of x, white spots form isolated islands, while at x nearly equal to unity, the isolated spots are black. We want to find the critical value of x at which a noninterrupted path appears (or disappears) across the whole system, going only through white regions.

Likewise, we can also formulate the corresponding 3D problem, filling a large volume with a white and a black substances. Then we have to vary the fraction of volume occupied by one of these substances until percolation sets in.

How to Construct a Random Function **

This new problem is not a lattice problem. Let us analyze its mathematical formulation in the plane case. We have to define on the whole plane a *random function* $V(X, Y)$, where X and Y are coordinates. In the Flood problem this function is the altitude above sea level of the point on the earth's surface with the given coordinates. We shall confine ourselves only to the case of *Gaussian random functions* (named after the great mathematician Karl Friedrich Gauss).

The simplest method of constructing a Gaussian random function is to ascribe to each point in space a random number not related in any way to a neighboring random number. Such random functions were given the name "white noise". The values it takes at neighboring points differ abruptly, that is, this function is discontinuous. To obtain a continuous function, the white noise has to be "smoothed". This procedure consists in ascribing to each point of space a quantity equal to the average of the values that the "white noise" function assumes in a certain region around this point. The continuous random function is then formed by these values. The function varies only slightly from this point to the neighboring point because of the small size of the spatial region over which the white noise is averaged. Let us denote by r_0 the size of the region over which averaging is carried out. This size is referred to as the *correlation radius* of a random function. Its main property is that the function changes only slightly when the argument of the function changes by an increment small compared with r_0.

The quantities V can be described by a distribution function that we denote by $f(V)$. By definition, the probability for the function $V(X, Y)$ to have at a point taken randomly in space a value within a small interval from V_1 to $V_1 + \Delta V$ equals $f(V_1)\,\Delta V$.

The function $V(X, Y)$ can always be constructed in such a way that, averaged over all points of a plane, it gives zero. To achieve this, we need to use in the construction of the "white noise" function the random numbers distributed symmetrically with respect to zero. It can be shown that the function $V(X, Y)$ obtained from such "white noise" has the Gaussian distribution (see Chapter 2) of the type

$$f(V) = \frac{1}{\delta\sqrt{2\pi}} \exp\left(-\frac{V^2}{2\delta^2}\right) \qquad (1)$$

Note that the probability density (1) is symmetric relative to the positive and negative values of V. In "mountain lingo" this means that peaks and valleys are encountered with equal probability.

Now we are ready to enunciate the percolation problem. The contours shown in Fig. 40 are determined by the condition $V(X, Y) = \text{const}$. In order to prescribe the water level, we need to introduce a quantity t that can vary from $-\infty$ to $+\infty$. The regions on the plane where $V(X, Y) < t$ will be referred to as white (covered with water), and those where $V(X, Y) > t$ as black (protruding above water). The thick contours in Fig. 40 correspond to $V(X, Y) = t$.

The percolation level is defined as the critical value t_c at which white regions form paths going uninterruptedly across the whole system. It is also possible to speak of the critical fraction x_c of space filled with white regions at the moment when percolation sets in. This fraction of space equals the probability for a continuous random variable V to assume a value in the range $-\infty <$

$< V < t_c$. By virtue of the definition of distribution functions,

$$x_c = \int_{-\infty}^{t_c} f(V)\, dV \tag{2}$$

Formula (2) relates the critical fraction x_c of space to the percolation level t_c.

Analogy to the Site Problem **

Imagine that a plane lattice (its type is immaterial) with small period is superposed over the geographic map shown in Fig. 40. Let the lattice period be substantially smaller than the characteristic extensions of land and water regions. Let us "paint" the lattice sites in water-covered areas white, and those in land areas black. As in the familiar site problem, we again consider white sites connected if they are nearest neighbors. The critical fraction of area under water equals the critical fraction of white sites at which percolation through white sets in.

However, the new problem is not identical to the site problem. Remember Chapter 4 where the construction of a system of white and black sites has been described in detail. (Sites were labelled in that chapter as blocked and nonblocked.) It was an important aspect of that construction that each site became black or white as prescribed by a random-number generator, in no relation to the color of the sites in the nearest neighborhood. As a result, the white and black sites were thoroughly mixed. The situation is quite different in the new problem.

The lattice period being very small, black and white sites form large groups. Neighbors of a white site are almost certainly white, and those of a black site are most probably black. (The size of blocks is determined by the correlation radius r_0.)

It must be clear that to introduce a lattice into the percolation level problem is a purely formal operation. This is in fact a nonlattice problem, and its solutions must be independent of both the lattice period (provided it is sufficiently small) and lattice type. However, this method allows us to make use of a well-developed apparatus of lattice problems.

In Chapter 4 we gave an algorithm of the Monte Carlo solution of a site problem with a computer. This algorithm is entirely transferable to the percolation level problem. An array $K(X, Y)$ consisting of zeros and unities is generated by addressing a random function $V(X, Y)$ by the method described in Chapter 4. White sites correspond to unities, and black sites to zeros. Then percolation paths are searched for, and the critical fraction of space at which percolation sets in is determined by the same techniques.

Percolation Levels in Plane and Three-Dimensional Problems **

A plane problem has an exact solution if the properties of a random function $V(X, Y)$ are on the average symmetric with respect to $V = 0$. The Gaussian functions described above are among the functions with such properties.

In order to arrive at the exact solution, we must use the symmetric formulation of the percolation problem given in Chapter 5. As before, we refer to the regions in which $V(X, Y) < t$ as white, and to those in which $V(X, Y) > t$ as black. In addition to percolation level through white regions, t_c, we can introduce percolation level through black regions, t'_c. The further proof goes by the following steps:

1. By virtue of the symmetry of the function $V(X, Y)$, we have $t_c = -t'_c$. Indeed, replace V by $-V$ at each point of the plane. This will give a function $V' = -V$ that on the average has the same properties, so that percolation levels calculated by V' must be equal to those calculated by the original function V. The inequality $V < t_c$ yields the inequality $-V > -t_c$, that is, $V' > -t_c$. At $t = t_c$ the regions that are white with respect to the function V ($V < t_c$) form an infinite cluster. These same regions of space are black with respect to the function V' and $t = -t_c$. Indeed, here $V' > -t_c$. Therefore, the value $t = -t_c$ is percolation level through black regions for the function V'. But it has already been said that the functions V and V' should have identical percolation levels. Consequently, percolation level through black regions for the function V is $t'_c = -t_c$.

2. If $t_c < 0$, then, as t increases, first percolation through white regions appears (at $t = t_c < 0$) and then (at $t = -t_c > 0$) percolation through black regions disappears. In the range $t_c < t < -t_c$ there is percolation both through white and black regions. If $t_c > 0$, first percolation through black regions disappears, and

only then percolation through white regions sets in. And in the range $-t_c < t < t_c$ percolation is impossible.

3. In plane problems, percolation through black regions excludes percolation through white regions, and vice versa. Indeed, if we can sail across a mountain ridge from west to east, it means that we cannot cross it by land from north to south. This excludes the case $t_c < 0$. On the other hand, the absence of percolation through white regions necessarily means the presence of southward percolation through black regions. (You can make sure of that by analyzing pictures like Fig. 40.) Therefore, the case $t_c > 0$ is also excluded. This leaves us with the single possibility: $t_c = 0$. This is the sought result: *percolation level stands at zero.*

Formula (2) makes it possible to calculate the critical fraction x_c of area. As follows from the normalization condition for distribution function (see formula (1) of Chapter 2), at $t_c = \infty$ the right-hand side of formula (2) equals unity. The symmetric nature of function $f(y)$ implies that at $t_c = 0$ the fraction of area $x_c = 0.5$.

In three-dimensional cases, eastward percolation through white regions does not preclude southward percolation through black regions because percolation channels can be easily uncoupled. (Remember highway bypasses arranged at different levels.) For this reason, in the 3D case, $t_c < 0$ and, by virtue of formula (2), $x_c < 0.5$. Monte Carlo calculations by the above-described procedure for Gaussian random functions demonstrated that the 3D critical fraction $x_c = 0.16 \pm 0.01$.

The quantity x_c can be approximately evaluated by the method we used to evaluate percolation thresholds of the site problem in Chapter 6. The method consists in surrounding each site by a sphere (or circle in the plane case) of radius equal to the distance to the nearest neighbor. A sphere around a white site is regarded as white, and a sphere around a black site is regarded as black. It was found that percolation through white spheres in contact appears when the fraction of space filled with these spheres is approximately the same for all lattices. It is natural to assume that this fraction must not be very different from the value x_c that we encounter in the percolation level problem. According to formulas (3) and (4) of Chapter 6, in the 2D case the fraction of area occupied by white circles is approximately 0.5, and in the 3D case the fraction of volume occupied by white spheres is approximately 0.16. The two estimates thus coincide strikingly well with the solution of the percolation level problem. We can expect that $x_c = 0.16$ is a fairly good estimate for non-Gaussian random functions.

Impurity Compensation in Semiconductors

Percolation level plays a very important role in the theory of extrinsic semiconductors. Assume that a semiconductor was doped by both donor and acceptor impurities in identical amounts. Donor impurity atoms have an extra electron on the outer electron shell, while, contrary to this, acceptor impurity atoms lack one electron. As a result, donors eagerly donate their extra electrons to acceptors and thereby become positively charged. Acceptors accept electrons and become negative. (This phenomenon is called *impurity compensation*.) Impurities in semiconductors being distributed chaotically, a random system of positively and negatively located charges is formed. Each charge produces an electric potential $\pm e/\varepsilon r$, where e is the absolute magnitude of the electron charge, ε is the dielectric constant, and r is the distance from the charge. The sign of the potential depends on the sign of the charge.

The potential of any point in space is the sum of the potentials produced by all impurity atoms. The impurities being distributed randomly through the semiconductor, their potential is also a random function.

If the concentration of donors is slightly higher than that of acceptors, some electrons remain on the donors. If the donor-electron bonding energy is relatively low, the thermal motion of atoms easily ionizes the donors. In principle, the liberated electrons can participate in current transfer, but this transfer is inhibited by an electric potential produced by charged impurity atoms within the semiconductor. The product of electric potential by the electron charge is the potential energy of interaction between an electron and the electric field of impurities. Potential energy essentially changes the character of motion of electrons.

Motion of a Particle with Nonzero Potential Energy

The total energy E of a particle consists of its kinetic energy $mv^2/2$, where m is the mass of the particle and v is its velocity, and potential energy V (r) which is a function of coordinates of the point at which the particle is located:

$$E = \frac{mv^2}{2} + V(\mathbf{r}) \tag{3}$$

where **r** is a vector from the origin of coordinates to the location of the particle.

The trajectory of a particle is described by a function $\mathbf{r}(t)$, where t is time. The fundamental principle of mechanics is that the total energy E of a particle does not change in the course of motion. This means that a change in coordinates of the particle entails a change in its velocity so as to compensate for the change in the potential energy $V(\mathbf{r})$. The law of energy conservation and the fact that the kinetic energy $mv^2/2$ must be a positive quantity impose important restrictions on the motion of the particle.

Let us assume for the sake of simplicity that V is a function of only one coordinate, namely, X, and that velocity is also directed only along the X-axis. Let $V(X)$ have the form shown in Fig. 41. The type of motion of the particle is determined by its total energy. If the total energy is E_1, the kinetic energy is positive only in the range $X_1 < X < X_1'$. Points X_1 and X_1' are the bounds of motion. Velocity vanishes at these points, hence the term *turning points*. A particle with energy E_1 is trapped into the *potential well* between the turning points and cannot leave this well. Its range of motion is bounded on both sides. A particle with energy E_2 cannot penetrate deeper on the left than point X_2, but its range of motion is not bounded on the right.

The motion of a particle can thus be bounded or unbounded depending on the relative values of the total and potential energies.

Note that quantum mechanics allows a particle to penetrate into regions with a negative kinetic energy. However, if V (**r**) is a sufficiently smooth function (and in semiconductors with high concentration of compensated impurities it is always smooth), this penetration plays only a minor role.

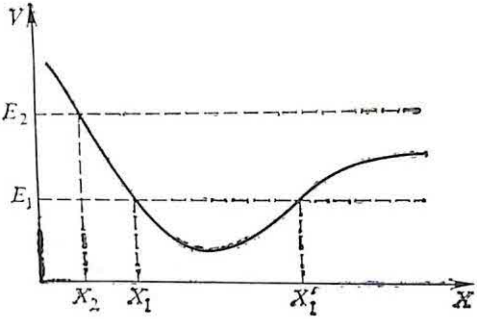


Fig. 41. Potential energy V of an electron as a function of the X-coordinate (the solid curve).

Motion of an Electron in the Field of Impurities

The potential energy of an electron, stemming from the potentials of randomly distributed impurity atoms, is a random function of coordinates. If the total energy E of the electron is low, it can move only in a bounded region of space everywhere surrounded by turning points. If the energy is sufficiently high, the electron is allowed to move throughout the whole space. This is the necessary condition for participating in the transfer of electric charge (in electric current).

The reader must understand the difference between the one-dimensional problem discussed in the preceding section and two-dimensional or three-dimensional problems. If a particle can move only along a straight line, then the motion can be unbounded in both directions only if the energy of the particle is greater than all the values assumed by the potential energy along this straight line. Hence, a stringent condition $E > V_{\max}$ must be satisfied ($V_{\max}$ is the maximum value of potential energy).

No such condition is required in the two- and three-dimensional cases. A particle can *bypass* the regions in which its motion is forbidden. It is only necessary that the regions where the motion is allowed form a system of "lakes and canals" through which the particle could travel to infinity in an infinite system.

Now it is obvious that the problem that emerges is one of finding percolation level. Let $V(\mathbf{r})$ be a random function describing the electron potential energy. We fix the total energy E and refer to spatial regions where $E > V(\mathbf{r})$ (the positive kinetic energy) as white, and to those where $E < V(\mathbf{r})$ as black. We want to find percolation level, that is, the critical value E_c at which percolation through white regions sets in.

Only those electrons whose energy exceeds E_c are free and take part in current transfer (sometimes this energy is called the *mobility edge*).

At low temperatures, E_c is substantially greater than the energy of thermal atomic motion kT. Consequently, the probability for an electron to reach the level of E_c is low. Correspondingly, the concentration of electrons capable of charge transfer is small. As temperature increases, this concentration grows steeply, thereby increasing the electric conductivity of the system. The percolation threshold E_c thus determines the temperature dependence of the electric conductivity of impurity-compensated semiconductors.

At very low temperatures, electrons accumulate in the deepest potential wells and cannot take part in charge transfer. Consequently, at low temperatures, impurity-compensated semiconductors turn into insulators.

It was mentioned in Chapter 8 that at a sufficiently high concentration of impurities of one sort (e.g. donors), the semiconductor's conduction becomes metallic, only marginally depending on temperature down to absolute zero (the Mott transition). Impurity compensation (e.g. by adding acceptors) enhances the random potential energy and switches off the metallic conduction. A detailed theory of this phenomenon has been developed on the basis of percolation level concepts.

Part III

Critical Behavior of Various Quantities Near Percolation Threshold. Infinite Cluster Geometry.

In this part we discuss those aspects of percolation theory that appear as the most interesting from physics' standpoint: the behavior of various quantities in the immediate neighborhood of percolation threshold. It was said in the previous parts that such physical quantities as the spontaneous magnetization of a doped ferromagnetic or the electric conductivity of a network with blocked sites vanish at the threshold point. In the present part we discuss the laws that describe their behavior in the vicinity of percolation threshold. The derivation of these laws requires that the geometrical properties of infinite clusters be understood.

Chapter 11**

The Bethe Lattice

It was shown in Chapter 5 that exact solutions can be found in some plane percolation threshold problems. However, it was never said that it was also possible to find the function $P(x)$, that is, the probability for a site to belong to an infinite cluster. No exact expressions for this function (or for the electric conductivity of a network) are known at present either for plane or for three-dimensional problems. The only exception is the Bethe lattice which, as we shall show below, must be classified as a lattice in the space of infinitely high dimensionality. In what follows we will give the formulation and solution of the site problem on the Bethe lattice.

Rumors

"What nonsensical stories they do spread about the town. What are things coming to when you can hardly turn round before there is some scandal going about you, and not a word of sense in it either..."*—in these words Nikolai Gogol described how a preposterous gossip spread by two ladies ruined Chichikov's promising fraud. "This enterprise they contrived to carry out in just a trifle over half an hour. The town was positively stirred up; everything was in a ferment; and if there were but anybody that could make out anything!"**

And indeed, rumors do spread unbelievably fast. But this speed ceases to be surprising if we analyze the mathematical model suggested below.

---

[*] *Dead Souls*, a poem by N. Gogol. Chatto & Windus, London, 1922; translated from the Russian by Constance Garnett.

[**] *Dead Souls*, a poem by N. Gogol. The Heritage Press, New York, 1944; translated from the Russian by B. G. Guerney.

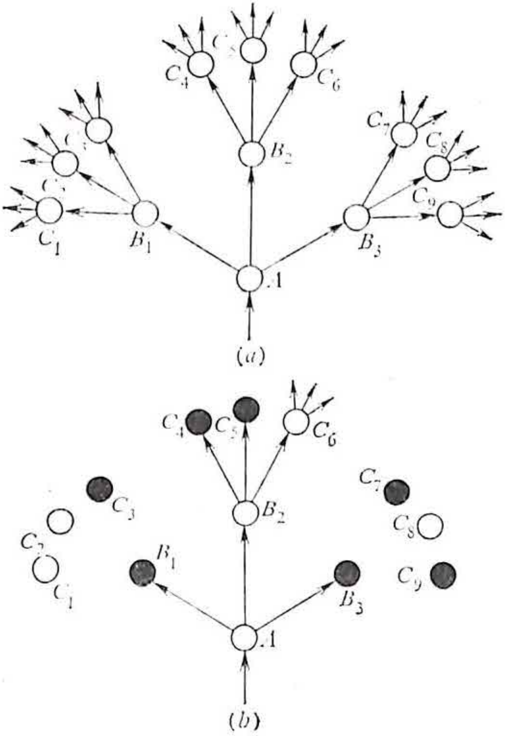


Fig. 42. The Bethe lattice for $q = 3$. The open circles represent people of category O, and the black circles represent those of category T.

The model is shown in Fig. 42*a*. Assume that a "lady agreeable in every respect", denoted in the figure by circle A, passed the news to three of her acquaintances, B_1, B_2, B_3. Each of these three passed the message on to three of her (or his) rumor-mongering acquaintances, so that this "second-hand" information was received by nine people marked by circles C. Each of these nine people again transmitted information to their three contacts, making the news known to another 27 people. It is readily found that the "tenth-hand" information reaches $3^{10} = 59\,049$ people! Assuming that it takes each gossip twenty minutes to pass the information to his (her) three listeners, we find that the whole chain takes 200 min = 3 h 20 min.

Of course, this model strongly simplifies the real process. It assumes that all people have an identical number of acquaintances. Furthermore, it assumes that each person receives information from only one informant. This means that only one line enters each circle (see Fig. 42*a*). By virtue of this property, the model resembles a tree that branches out infinitely in all directions. Each circle can be regarded as a trunk of its own tree, and the trees that grew from, for instance, circles B_1, B_2, B_3 have no common circles. The same can be said about the trees stemming from circles C, and so on.

In scientific literature, this model is called just that: a *tree*. It is also called the *Bethe lattice*, after the famous physicist Hans Bethe. The circles shown in Fig. 42*a* are the sites of this lattice. The number of lines originating from each site of the Bethe lattice can be arbitrary (but identical for all sites). Let us denote this number by q. Figure 42*a* shows the lattice for $q = 3$.

Now recall that most people have their own point of view and do not participate in spreading rumors. Let us divide all people into two categories: category O shown by the open circles consists of people who transmit the received information to the next trio (Fig. 42*b*). These circles give rise to q arrows. Category T shown by the black circles consists of people who do not participate in spreading the rumors. No arrows originate from the black circles (see Fig. 42*b*).

The introduction of the black circles strongly affects the propagation of a rumor. Consider the configuration shown in Fig. 42*b*. Among the three circles B_1, B_2, B_3, only one is white and transmits the rumor further on. Circles C_1, C_2, C_3 would be glad to gossip but received nothing from B_1 and B_3. Among C_4, C_5, C_6, only C_6 belongs to O. Therefore, instead of nine recipients of the second-hand information, only three were reached by the rumor, and only C_6 will transmit it along the line.

Imagine that the system under discussion is not bounded and thus has an infinite number of circles. Then the following question can be posed: Will a rumor originating from point A die out after a finite number of transfers, or will it stray to an infinite distance from A and become known to an infinite number of persons in an infinite system? We see from Fig. 42*b* that this depends on the relative number of open and black circles and on configurations that arise around a site.

In fact, we speak of a site problem of percolation theory, only formulated on the Bethe lattice. Let the fraction of people belonging to O be x. This means that a person selected at random will be found to belong to category O with probability x and to category T with probability

$1 - x$. The question we have to answer is as follows: What is the probability $P(x)$ for a rumor transmitted to a randomly selected person to become known to an infinite number of people? Obviously, this probability is zero if x is small, but becomes greater than zero beginning with a certain critical value $x = x_c$.

Solution of the Site Problem on the Bethe Lattice

It will be convenient to introduce, instead of $P(x)$, the probability for a rumor transmitted to a randomly chosen person not to become known to an infinite number of persons. Let us denote this probability by $Q(x)$. Obviously,

$$Q(x) = 1 - P(x) \tag{1}$$

because these two events form a complete system.

An algebraic equation can be written for $Q(x)$. The following line of reasoning must be used. The spreading of a rumor can be stopped for two incompatible reasons. The first is that the person chosen at random belongs to category T. The second reason is that although this person belongs to category O and passes the rumor to q people, all channels leading from these people will be cut at different stages. The probability Q is thus the sum of the probabilities of two incompatible events. The probability for a randomly selected person to belong to category T is $1 - x$. The probability that the person belongs to category O but that the rumor propagation will be interrupted at later stages is denoted, for the time being, by W'. Then

$$Q = 1 - x + W'$$

Now let us take a look at W'. The outcome that it represents is the result of two simultaneously realized events: (i) the person chosen randomly belongs to O (the probability of this event is x), (ii) all q channels leading from the acquaintances of the person chosen at random interrupt at some step. Obviously, these two events are independent. Hence, the probability W' equals the product of the probabilities: $W' = xW(x)$, so that

$$Q = 1 - x + xW(x) \tag{2}$$

where $W(x)$ is the probability for all q channels to be interrupted at some step (of course, this can happen at different steps in different channels).

Let us consider one of the q channels that originate from one of the acquaintances of the person chosen at random. The event consisting in this channel being interrupted at some step is equivalent to the statement that the rumor communicated to this acquaintance does not reach an infinitely large number of people. By definition, the] probability of this event is $Q(x)$.

It is very important for what follows that the trees rooted at q acquaintances of a randomly selected person have no common circles. Hence, a certain configuration of open and black circles found in one tree does not affect at all the probability of any possible configuration of circles in other trees. (Obviously, this last statement would be incorrect if the trees had common circles.) Consequently, the events consisting in the interruption of the rumor in one or another channel are independent.

The probability for all q channels to be interrupted is therefore equal to the product of the probabilities for each of the q channels to be interrupted:

$$W(x) = [Q(x)]^q \tag{3}$$

Substituting formula (3) into (2), we obtain an equation for $Q(x)$:

$$Q(x) = 1 - x + x[Q(x)]^q \tag{4}$$

Note that the decisive factor that enables us to reduce the problem to algebraic equation (4) is the independence of channels. This property is an exclusive property of the Bethe lattice, so that the method we use is not successful when applied to usual lattices, although it is often employed to obtain an approximate solution.

Let us analyze equation (4). It is meaningful for all x within the interval $0 \leqslant x \leqslant 1$. Rewriting it in terms of $P(x) = 1 - Q(x)$, we find

$$[1 - P(x)]^q x + P(x) - x = 0 \tag{5}$$

One of the solutions of (5) is $P(x) = 0$ for all x, although for $q > 1$ equation (5) is nonlinear and has other solutions. Thus, $P(1) = 1$ is also a solution for $x = 1$, and this second solution is physically meaningful because the probability P must be unity and not zero if all circles are open.

For $q = 2$ the solutions of equation (5) are easily found. There are two solutions in this case: $P(x) = 0$ and $P(x) = 2 - 1/x$. If $x >$

$> 1/2$, the second solution has a physical meaning. If $x < 1/2$, the solution becomes negative and thus meaningless.

Therefore, for $q = 2$ we have the following solution:

$$P(x) = \begin{cases} 0 & \text{for } 0 \leqslant x < 1/2 \\ 2 - 1/x & \text{for } 1/2 < x \leqslant 1 \end{cases} \tag{6}$$

In this case the percolation threshold x_c equals 1/2.

A similar solution exists for all $q > 1$, although the percolation threshold x_c is a function of q. In the general case it is possible to find x_c and the form of $P(x)$ for x close to x_c, assuming from the start that $P(x) \ll 1$ (this is always true in the neighborhood of percolation threshold). The term $(1 - P)^q$ in equation (5) can be expanded via the binomial theorem:

$$(1 - P)^q = 1 - qP + \frac{q(q-1)}{2} P^2 - \ldots \tag{7}$$

Each subsequent term here is much less than the preceding term because $P \ll 1$. Therefore, we substitute (7) into equation (5), assuming that only the three terms written above are significant on the right-hand side of (7). This gives

$$\frac{q(q-1)}{2} xP^2 = qPx - P$$

Assuming $P \neq 0$, we divide both sides of this equality by P and obtain

$$P = \frac{(x - 1/q) \cdot 2}{x(q-1)} \tag{8}$$

Formula (8) coincides with (6) for $q = 2$. This solution vanishes at $x = 1/q$, which implies that $x_c = 1/q$. Solution (8) is meaningful if $q > 1$, $x > 1/q$, and only if x is very close to $1/q$. Therefore, we can set $x = 1/q$ in the denominator of (8). Finally, we obtain

$$P(x) = \frac{(x - 1/q) \cdot 2q}{q - 1} \tag{9}$$

Formula (9) describes the probability $P(x)$ in the vicinity of percolation threshold.

Discussion

If $q = 1$, the function $P(x) = 0$ for all x within the range $0 \leqslant x < 1$, and equation (5) has no other solutions within this range. If $q = 1$ and $x = 1$, equation (5) holds for any value of P.

If $q = 1$, the Bethe lattice transforms into a linear chain of lattice sites. And no matter how small number of black sites interrupts percolation through white in such a chain. It is therefore natural that whatever the value of x within the range $0 \leqslant x < 1$, the interruption of the rumor propagation is unavoidable, that is, $P(x) = 0$. In a linear chain $x_c = 1$.

We have shown in the preceding section that in the general case $x_c = 1/q$. This result could be predicted in advance. Each person in the problem under discussion transmits the rumor to q of her (his) acquaintances. The mean number of people of category O among these acquaintances equals qx. Hence, each transfer of information on the average creates qx sources instead of one. The quantity qx is therefore its *branching ratio*.

In order for the process to continue, the branching ratio must be greater than unity. Hence, the critical concentration x_c is obtained from the condition $qx_c = 1$, that is, $x_c = 1/q$.

Recall that the condition necessary to sustain the uranium fission chain reaction is written in the same form. In fact, the process of rumor spreading is also a chain reaction and is described in the same terms as a nuclear explosion.

It is of interest to compare the obtained value of x_c with the results derived for lattice problems in spaces with a greater dimensionality. Percolation threshold of the site problem was approximately computed for the so-called *hyperlattices*. These are lattices of the same type as the square and simple cubic ones but in a space with a greater number of dimensions. The coordination number z (the number of nearest neighbors) is found in such lattices from the formula $z = 2d$, where d is the dimensionality of space (it gives $z = 4$ for $d = 2$, and $z = 6$ for $d = 3$). The percolation threshold x_s was computed for $d = 4, 5, 6$. It was found that the formula

$$x_s = \left(1 + \frac{6.3}{d^2}\right) \frac{1}{z-1} \qquad (10)$$

provides a good fit to the obtained results.

The second term in parentheses can be neglected if d is sufficiently large, so that $x_s = 1/(z - 1)$. But recall that $z - 1 = q$ for the Bethe lattice. Indeed, only one bond enters each site of this lattice, and q bonds originate from it.

We see that percolation threshold of the Bethe lattice ($x_c = 1/q$) is the same as that of a hyperlattice with a very large dimensionality. Hence, the Bethe lattice as if represents an infinite-dimensional space.

The Bethe lattice is the only system for which it proved possible to find the exact form of $P(x)$ at percolation threshold. It was found (see (9)) that in this case $P(x)$ tends to zero *linearly* as $x \to x_c$: $P \propto (x - x_c)$. We shall see later that this is a specific property of the Bethe lattice as well as of all lattices of high dimensionality.

Exercise

1. Analyze the bond problem on the Bethe lattice. Consider all sites to be identical, and the bonds to be intact or broken. Let the fraction of intact bonds be x. Find the function $P(x)$ defined as in the text.

Chapter 12

Structure of Infinite Clusters

The Shklovskii-de Gennes Model

Now we shall consider, for the sake of definiteness, the site problem and assume that the concentration of nonblocked sites is slightly above the threshold value, so that an infinite cluster exists. It consists of infinite strings (chains) of interconnected sites. If all connected sites of an infinite cluster are joined by straight segments, we obtain a set of intersecting broken lines (see Fig. 15 showing one such line).

By definition, the structure of an infinite cluster is its geometry on a scale much greater than the lattice period. On this scale, the sharp knees occurring at individual lattice sites are not resolved to the eye, and strings appear as smoothly bent curves.

Figure 43 shows a small fragment of an infinite cluster. The cluster does not end at points A and B but stretches leftwards and rightwards to infinity. Now let us introduce the following classification of points and lines in an infinite cluster: the elements of an infinite cluster belong to either its *backbone* or *dead-ends*.

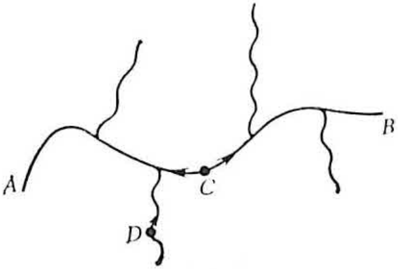


Fig. 43. Fragment of an infinite cluster with dead-ends

A point is said to belong to the backbone of an infinite cluster if there are at least two paths that emerge from this point in opposite directions and lead to infinity. Point C in Fig. 43 is one of such points. Going to the left or to the right of this point, we can move infinitely far from it. If only one path leaving a point leads to infinity, this point belongs to a dead-end. For instance, only the motion upwards from point D in Fig. 43 leads to infinity. The downward motion ends at a cul-de-sac. Point D is thus said to lie on a dead-end.

Let us imagine that all dead-ends are deleted; now we can try to discern the structure of the backbone of an infinite cluster. The simplest backbone model was suggested independently by the Soviet physicist B.I. Shklovskii and the French physicist P.G. de Gennes. In a plane problem, this model resembles a very large, old, and fairly worn-out fishnet. The net lost its regular periodicity, its cords have slack, some nots are broken, while others slipped off to wrong locations, but nevertheless it's still a net "on the average" (Fig. 44).

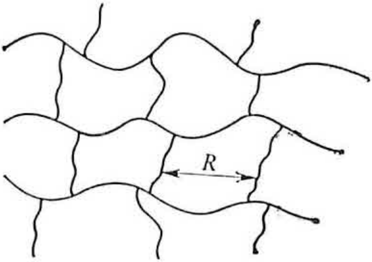


Fig. 44. Backbone of an infinite cluster.

The characteristic linear size R of a cell of this net is called the *correlation radius* of an infinite cluster. It grows drastically as we approach percolation threshold:

$$R = \frac{l}{|x - x_c|^{\nu}} \tag{1}$$

where l is a length that in order of magnitude equals the lattice period, and ν is a positive quantity called the *exponent of correlation radius.* The network thus becomes less and less dense as percolation threshold is approached.

A correlation radius approaching infinity is the general property of all critical phenomena.

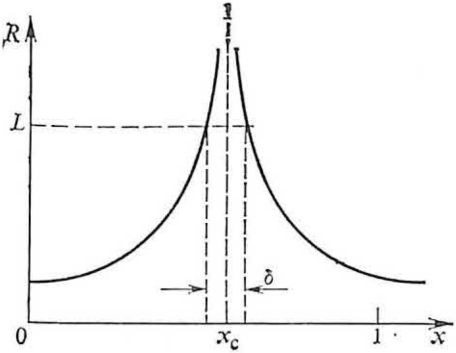


Fig. 45. Correlation radius as a function of x. The graph shows the width δ of the critical region for the $L \times L$ square (see the next section).

The power-law behavior (1) is not a rigorously proved property; nevertheless, it forms the basis for the modern theories of critical phenomena and appears to be well supported by experimental data.

A correlation radius remains meaningful at $x < x_c$, that is, below the threshold. In this region it describes the *maximum size of finite clusters.* If $x \to x_c$ on the side of smaller values $(x < x_c)$, a correlation radius also tends to infinity as in (1). This means that when percolation threshold is approached from below, finite clusters grow infinitely and merge into an infinite cluster at $x = x_c$. The dependence of R on x thus has a shape schematically traced in Fig. 45.

For three-dimensional problems the Shklovskii-de Gennes model is formulated similarly. It resembles a badly damaged wire skeleton of a three-dimensional lattice whose mean cell size is given by formula (1). But it must be borne in mind that the numerical values of the exponents of correlation radius are different for plane and 3D problems.

Let us consider now the corollaries derived from the network-like structure of infinite clusters.

Role of the System's Size

We emphasized in Chapters 1, 2, and 3 that the concept of percolation threshold is truly meaningful only in an infinite system. In a finite system, percolation threshold varies among samples, that is, it is a random variable. However, the values assumed by this random variable fall, with very high probability, into a certain range of width $\delta(\mathcal{N})$ called the *critical region*. As the number of sites in the system grows, the width of this region undergoes a power-law decrease (see formula (8) of Chapter 1), so that percolation threshold acquires a clear-cut sense as $\mathcal{N} \to \infty$ and converts from a random variable into a certain quantity.

This information was practically quoted, with no attempt to derive it, in the opening chapters of the book. The concept of correlation radius makes it possible to understand these concepts and to derive formula (8) of Chapter 1.

For definiteness, let us discuss the wire-mesh experiment on a grid $L \times L$ in size, the schematic of this experiment being shown in Fig. 1. Assume that numerous experiments using various random sequences of blocked sites have been conducted, and the result is a set of percolation threshold values. Recall that the configurations of blocked sites are very dissimilar in different experimental runs.

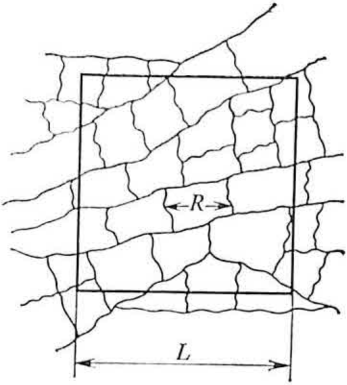


Fig. 46. Square superposed on an infinite network, with $R \ll L$ and $x > x_c$.

It is convenient to pursue the following line of reasoning. Imagine an *infinite* network (a wire mesh), with a prescribed bed fraction x of non-blocked sites. Imagine now that an $L \times L$ square is superposed on the mesh at different places, and percolation is analyzed from the left- to the right-hand side of this square through the non-blocked sites that fell within the square (Fig. 46).

Placing the square over the different parts of the infinite wire mesh, it is possible to sample one by one the results of successive experiments with a finite wire mesh.

Percolation sets in in an infinite network exactly at $x = x_c$, but as we shall see below, this does not mean at all that percolation is always realized in an $L \times L$ square when $x > x_c$.

If $x > x_c$, an infinite cluster exists in the infinite system. Let us picture its backbone as a fishnet (Fig. 44). The parameter extremely important hereafter is the ratio of correlation radius R to the square size L. First we assume that L is much greater than R. Then (see Fig. 46) the square covers a large number of unit cells of the infinite cluster netting that ensures percolation between the opposite sides of the square. These cells may be different in size, and the netting of the infinite cluster may contain large holes, but the probability for the cluster to contain a square-size hole is negligibly small if on the average the square contains very many cells. The following conclusion is therefore drawn:

> If $x > x_c$, percolation threshold of a square cannot lie within the range of x satisfying the strong inequality $L \gg R(x)$. This region must lie above the threshold. (I)

According to formula (1), a correlation radius grows infinitely when x tends to x_c, so that at some value of x the radius inevitably becomes comparable with L. Now we cannot say anything definite about percolation across this square.

Everything depends on the specificities of the configuration of blocked sites within the square.

Now let $x < x_c$, with the correlation radius much smaller than L. At $x < x_c$ the correlation radius is the maximum length of finite clusters. If $R \ll L$, no cluster exists long enough to connect the opposite sides of the square. Hence, another definite conclusion is made:

If $x < x_c$, percolation threshold of a square also cannot lie within the range of x satisfying the strong inequality $L \gg R\ (x)$. This region must lie below the threshold. (II)

If $x < x_c$ but x is very close to x_c, the correlation radius exceeds L. In this case we cannot say anything definite about percolation across the square. An infinitely large system contains finite clusters whose size is greater than L, but these clusters have holes of a similar size, so that everything depends on a specific configuration of blocked sites within the square.

Now we can evaluate the size of the critical region in which the values of percolation threshold for an $L \times L$ square can lie. According to the corollaries (I) and (II), this region must obey the condition $L \leqslant R$. Figure 45 shows that this region is the narrower and squeezes up to percolation threshold for an infinite system the closer, the larger L is. The width δ of the region obeys the condition $R\ (\delta) = L$. From formula (1) we find $l/\delta^{\nu} = L$, or

$$\delta = \left(\frac{l}{L}\right)^{1/\nu} \qquad (2)$$

Percolation thresholds of $L \times L$ squares are distributed uniformly within the critical region, that is, for $|x - x_c| \ll \delta$ (see Fig. 5 where the distribution function of percolation thresholds is plotted). Nothing singles out the point $x = x_c$ within this region. Indeed, this is the point at which percolation sets in in an infinite system. But it is impossible to establish whether percolation is achieved or not if we work with finite-size squares. If $L < R$, it is impossible to establish, by placing the square over different regions of an infinite network, whether the network contains only finite clusters or whether these clusters have already merged and formed an infinite cluster. *A study of percolation across a finite-size square only allows us to determine the width of the critical region.*

In this section we discussed only plane problems. In fact, the arguments completely transfer to three-dimensional problems. Formula (2) determines the width of the critical region in 3D problems equally well. A slight difference appears if the width δ is written not in terms of the system's size L but of the total number of sites, $\mathcal{N}$. The point is that $\mathcal{N} = (L/a)^d$, where a is the lattice period, and d is the dimensionality of space. Consequently, we have by (2)

$$\delta(\mathcal{N}) = \frac{C}{\mathcal{N}^{1/\nu d}} \qquad (3)$$

where C is a numerical coefficient that cannot be found by equally simple reasoning. In plane cases ($d = 2$), formula (3) coincides with formula (8) of Chapter 1. It was by means of this formula that the exponent of correlation radius of the plane problem was established for the first time from the dependence $\delta\ (\mathcal{N})$ determined numerically on a computer. It was found that $\nu_2 = 1.33$. (Here and below the subscript 2 indicates that we deal with the exponent of a two-dimensional system.) The exponent ν of three-dimensional problems is different: $\nu_3 = 0.8\text{-}0.9$. (The subscript 3 denotes that the exponent refers to 3D problems.)

Electric Conduction Near Percolation Threshold

Let us consider specifically two- or three-dimensional networks with blocked sites. As we said in the opening chapters, such networks are electrically conductive at $x > x_c$, but at the percolation threshold x_c their conductivity vanishes. Experimental data and the data obtained by computer simulation show that the specific electric conductivity of networks tends to zero in a power-law fashion,

$$\sigma\ (x) = \sigma_0\ (x - x_c)^t \tag{4}$$

where the factor σ_0 equals, by the order of magnitude, the specific electric conductivity of the network without blocked sites. The quantity t is called the *critical exponent of electric conduction*; it constitutes the subject of a very careful analysis, mostly by numerical computer simulation (e.g. one of the more recent computations involved a square lattice with 800 × 800 sites). It was established that $t_2 = 1.3$ in two-dimensional networks, and $t_3 = 1.6\text{-}1.7$ in three-dimensional networks, The network model of an infinite cluster makes it possible to derive formula (4) and to relate its exponent t to the exponent of correlation radius. Electric current flows only through an infinite cluster, namely, through its backbone. No current flows in dead-ends that couple to the backbone only at one end. If electric current were sufficiently strong, so as to heat the wire to a glow, the backbone of an infinite cluster would be visually discernable in the dark, as a rate of illuminated channels against a dark background. Far from threshold the whole network glows more or less uniformly, with the distance between glowing channels increasing as we approach the threshold; finally, the glow fades out when the threshold is reached: the current ceases to flow.

Let us calculate the specific electric conductivity of the backbone of an infinite cluster. It must be borne in mind that we cannot expect to calculate correctly the numerical factors. Our calculation will only give how σ depends on $x - x_c$. This dependence will not be altered if we replace an irregular, distorted network by an ideal network with period R.

First consider a plane case (see Fig. 46). The resistivity equals the resistance of a square with unit-length side. The number of wires intersecting this square is $1/R$, where R is the spacing between wires, given by formula (1). Let us denote the resistance of one wire with unit length by ρ_0. The circuit is the parallel connection of all wires. Therefore, the resistivity is

$$\rho = \frac{\rho_0}{(1/R)} = \rho_0 R \qquad (5)$$

and the electric conductivity is

$$\sigma = \rho^{-1} = \rho_0^{-1} R^{-1} \tag{6}$$

Substituting (1), we obtain

$$\sigma = \sigma_2 \, (x - x_c)^{\nu} \tag{7}$$

where $\sigma_2 = \rho_0^{-1} l^{-1}$.

In the three-dimensional cases we need to calculate the resistivity of a wire skeleton that models, for instance, the simple cubic lattice with period R (lattice type affects only the numerical coefficient). The resistivity equals the resistance of a cubic cell with unit-length edge. The number of wires connected in parallel passing through a face of such a cube equals $1/R^2$. The resistivity therefore equals

$$\rho = \frac{\rho_0}{(1/R)^2} = \rho_0 R^2 \tag{8}$$

and the electric conductivity is

$$\sigma = \rho_0^{-1} R^{-2} = \sigma_3 \, (x - x_c)^{2\nu} \tag{9}$$

where $\sigma_3 = \rho_0^{-1} l^{-2}$.

In order to avoid confusion, note that the electric conductivity σ has different dimensionalities in the two- and three-dimensional cases: it is measured in ohm^{-1} in the 2D case, and in $\text{ohm}^{-1} \cdot \text{cm}^{-1}$ in the 3D case.

The coefficients σ_2 and σ_3 correspond, by the order of magnitude, to the electric conductivities of two- and three-dimensional networks with no blocked sites. Indeed, as we see from formulas 6) and (9), the electric conductivity $\sigma(x)$ reduces to σ_2 or σ_3, respectively, if $R = l$, that is, if the network of an infinite cluster coincides with the initial network on which the problem has been formulated. The factor σ_0 in formula (4) is thus equal to σ_2 in the 2D case, and to σ_3 in the 3D case.

A comparison of formulas (6) and (7) with formula (4) shows that $t = \nu$ in the 2D case, and $t = 2\nu$ in the 3D case. Using $\nu_2 = 1.3$ and

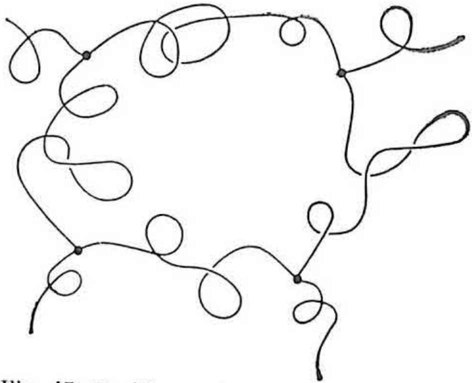


Fig. 47. Backbone of an infinite cluster, with "sinuosity" taken into account.

$\nu_3 = 0.8\text{-}0.9$, we find $t_2 = 1.3$ and $t_3 = 1.6\text{-}1.8$, quite close to the data given above. This agreement strongly supports the Shklovskii-de Gennes model.

The backbone model of an infinite cluster can be generalized as follows. Imagine that the wires forming the backbone are very sinuous (Fig. 47). The distance between their intersection points is, as before, $R(x)$ and is given by formula (1). However, if the wire segment between two intersection points is straightened out, its length will prove considerably greater than R. We denote this length by $\mathscr{L}$ and write it in the form

$$\mathscr{L} = \frac{l}{(x - x_c)^{\zeta}} \quad (x > x_c)$$

where $\zeta > \nu$. As $x \to x_c$, the ratio $\mathscr{L}/R = (x - x_c)^{-(\zeta - \nu)}$ tends to infinity.

The introduction of the length $\mathscr{L}$ is justified only if $\mathscr{L} \gg R$, that is, $\zeta > \nu$. However, it has been proved that $\zeta = 1$. Hence, the suggested generalization of the Shklovskii-de Gennes model (in fact, this generalization has been carried by the authors of the model) is meaningful only if $\nu < 1$. Since $\nu = 1.3$ in the 2D case, no "sinuosity" is expected there. But in the 3D case $\nu < 1$, and there is a good reason to believe that "sinuosity" indeed characterizes the backbone of an infinite cluster.

Exercise

1. Express the critical exponent t of electric conductivity in terms of the exponents ζ and ν in the 3D case, making use of the generalized model.

Function $P(x)$ Near Percolation Threshold. Role Played by Dead-Ends

The function $P(x)$, which is the fraction of sites belonging to an infinite cluster, vanishes at $x = x_c$, together with electric conduction. An analysis demonstrated that in the immediate vicinity of the threshold this function varies as

$$P(x) = D(x - x_c)^{\beta} \tag{10}$$

where D is a numerical coefficient of the order of unity, and β is another critical exponent. It has been found that $\beta_2 = 0.14$ (2D problems), and $\beta_3 = 0.4$ (3D problems). These results were mostly obtained in computer simulation.

All sites of an infinite cluster, both those belonging to the backbone and to dead-ends, contribute to $P(x)$. The infinite cluster model makes it possible to find which of these subsets has a greater number of sites. First assume that dead-ends are completely absent and calculate the contribution of the backbone of the infinite cluster to $P(x)$.

The number of sites belonging to the backbone per cell of an infinite cluster in the 2D case is of the order of R/a, where a is the lattice period (as in the preceding section, this is also an evaluation that is not meant to find numerical coefficients). The cell area is about R^2, and hence, the total number of all sites in a cell is about R^2/a^2. Therefore, the fraction of sites belonging to the backbone of an infinite cluster is

$$P_{\mathrm{bb}}(x) \sim \frac{a}{R} \sim (x - x_c)^{\nu} \tag{11}$$

Here the symbol $\sim$ stands for equality to within an order of magnitude (neglecting numerical coefficients of the order of unity).

In the 3D case the number of sites of the backbone per each cell of an infinite cluster is also R/a, but the total number of sites per cell is about $(R/a)^3$. Consequently, in the 3D case

$$P_{\mathrm{bb}}(x) \sim \left(\frac{a}{R}\right)^2 \sim (x - x_{\mathrm{c}})^{2\nu} \qquad (12)$$

Comparing formulas (11) and (12) with formulas (7) and (9), we obtain that the fraction of sites belonging to the backbone of an infinite cluster coincides, to within an order of magnitude, with the function $\sigma(x)/\sigma_0 = (x - x_{\mathrm{c}})^t$.

Comparing (11) and (12) with formula (10), we notice that

$$\frac{P_{\mathrm{bb}}(x)}{P(x)} \sim (x - x_{\mathrm{c}})^{\nu_2 - \beta_2}$$

in the 2D case, and

$$\frac{P_{\mathrm{bb}}(x)}{P(x)} \sim (x - x_{\mathrm{c}})^{2\nu_3 - \beta_3}$$

in the 3D case.

Recall that $\nu_2 = 1.3$ and $\nu_3 \approx 0.9$. Hence, $\nu_2 - \beta_2 = 1.2$ and $2\nu_3 - \beta_3 \approx 1.4$. Therefore, the ratio $P_{\mathrm{bb}}(x)/P(x)$ rapidly tends to zero as $x \to x_{\mathrm{c}}$ both in the 2D and 3D cases. This means that the sites forming the backbone of an infinite cluster comprise an infinitesimal fraction of the total number of sites belonging to the infinite cluster. *The "mass" of an infinite cluster predominantly resides in its dead-ends* and is completely useless from the standpoint of electric conduction. Consequently, in the vicinity of percolation threshold $\sigma(x)/\sigma_0 \ll P(x)$ (see Fig. 10). However, the spontaneous magnetization of a doped ferromagnetic in the neighborhood of percolation threshold is determined precisely by these dead-ends (see Chapter 3).

Universality of Critical Exponents

We have encountered three critical exponents that describe the behavior of various quantities in the neighborhood of percolation threshold, i.e. ν, t, and β. This behavior is also referred to as critical because the functions $R(x)$, $\sigma(x)$ and $P(x)$ have *singularities* at $x = x_c$. For instance, $R(x)$ tends to infinity, while $P(x)$ has a discontinuous first derivative (at x_c). The derivative is zero on the left of x_c and is infinitely large on its right. In the case of $\sigma(x)$, it is a second derivative that is discontinuous. Many other quantities with critical behavior are known in percolation theory, and correspondingly, many other critical exponents.

For each of the above-discussed critical exponents we gave two values: one for 2D and another for 3D problems. However, the list of relevant 2D and 3D problems would be extremely long. For instance, we know of three-dimensional site-, bond-, and sphere problems, the problem of percolation level in a random potential, and many others. What, then, were the exponents we discussed so far? Now we come to what seems to be the most exciting feature of percolation theory. According to modern concepts, *the critical exponents are identical for all problems in the space of a given dimensionality*. (The only exception among the problems mentioned in this book is the problem of directed percolation.) The statement of universality of critical exponents is rather a convention than a rigorously proved proposition. Nevertheless, numerous tests of this statement, carried out on computers, failed to refute it.

What are the physical reasons for the universality of exponents? Presumably, the exponents are determined by the structure of clusters in the vicinity of percolation threshold. The properties that play the principal role here are the geometrical properties of clusters that are felt at large distances (of the order of correlation radius). In the neighborhood of the threshold these distances become much greater than the lattice period (in lattice problems) or the radius of spheres (in sphere problems). Consequently, the geometry of clusters is independent of the type of lattice on which the problem is formulated. The structure of large clusters will remain unaltered even if a problem is given not on a lattice but on sites randomly distributed in space. But, of course, the dimensionality of space strongly affects the geometry of clusters because, for instance, it is much easier to ensure "bypassing" of curves in a three-dimensional space than in two dimensions. For these reasons, critical exponents do not depend on problem type, but do depend on space dimensionality.

It is interesting to note that critical exponents change with increasing dimensionality of space only up to dimensionality 6. Beginning with $d = 6$, exponents remain constant, and the exponent $\beta = 1$ as in the Bethe lattice. When $d \geqslant 6$, the critical exponent problem considerably simplifies and allows an exact solution.

Critical exponents thus possess a certain universality in contrast to percolation thresholds which strongly depend on the type of problem. This leads to a simple conclusion. If the results of a physical experiment are treated in terms of percolation theory, and the microscopic structure of the investigated system is not quite clear, the first characteristics to be compared with the theory must be critical exponents because they are almost invariant. This is the strategy one uses in analyzing experimental data on electric conduction in heterogeneous materials (see Chapter 9).

Percolation theory borrowed the idea of universality of critical exponents from the theory of second-order phase transitions (e.g. among second-order transitions are the transition of a metal from the superconducting to the normal state and the transition of a ferromagnetic to the nonferromagnetic state, both brought about by increasing the temperature of a sample). As in the vicinity of percolation threshold, large regions differing in properties from one another form close to the point of a second-order phase transition. The difference lies in that the boundaries of these regions are not "frozen" as they are in percolation theory, but change with time owing to thermal motion. The size of such regions is also called the correlation radius and also given by formula (1).

The theory of phase transition supplied another important idea as well: the *scaling hypothesis.* We shall formulate it in terms of percolation theory. Assume that different parts of a network with some blocked sites have been photographed at two values of x, namely, x_1 and x_2. Both these values are on one side of x_c and are close to it. Let, for instance, $x_1 - x_c > x_2 - x_c > 0$. At $x = x_1$ the correlation radius is $R_1 = l(x_1 - x_c)^{-\nu}$, and at $x = x_2$ it is $R_2 =$

$= l\,(x_2 - x_c)^{-\nu}$. These conditions imply that $R_1 < R_2$. The scaling states: if the photographs taken at $x = x_1$ are magnified by a factor R_2/R_1, they will not differ, "on the average", from the photographs taken at $x = x_2$. The photographs are assumed to be so coarse-grained that only large blocks show up, and individual sites and bonds are not resolved. The scaling hypothesis thus states that the large-scale geometry of the system is transformed on approaching percolation threshold in a self-similar manner, with all linear dimensions growing in proportion to the correlation radius.

Note that the Shklovskii-de Gennes model satisfies the scaling hypothesis, but that this hypothesis goes far beyond the model. It is neither restricted to the backbone of an infinite cluster nor does it assume that the network is subdivided into its backbone and dead-ends.

The mathematical formulation of scaling hypothesis makes it possible to find a relation between the critical exponents ν, β and a third exponent that we chose not to introduce here. Calculations show that this relation holds quite well.

Scaling concepts that were first introduced by the Soviet physicists A. Z. Patashinsky and V. L. Pokrovsky and by the American physicist L. Kadanoff form the basis of the modern theory of phase transitions and of percolation theory.

Scaling concepts brought about a number of new mathematical techniques for calculating critical exponents. A dramatic progress has been achieved in these methods in recent decade, so that certain perfection has been reached. At present, these methods help to calculate almost all important exponents. However, computations are so complicated that they could not be discussed in this book.

Chapter 13

Hopping Electric Conduction

The preceding chapters gave a detailed picture of how percolation theory is applied to calculate the electric conductivity of systems consisting of randomly coupled elements. Each element was either conducting or insulating, but the resistances of all conducting elements were assumed equal. Examples of such systems are networks with broken bonds or blocked sites, mixtures of metallic and dielectric spheres, and so forth.

Let us turn now to a different class of systems which are also composed of different elements, but the resistances of these elements can take on any values from extremely small to enormously large. It was found that the resistance of systems composed of a very large number of such elements can also be calculated in terms of percolation theory. The theory of hopping electric conduction of semiconductors, based on percolation ideology, was constructed in 1971. In this chapter we make acquaintance with hopping electric conduction and present its mathematical description.

Mechanism of Hopping Conduction

Consider an extrinsic semiconductor doped with, say, donor impurities. The dopant concentration will be considered small in comparison with the critical concentration N_c at which the Mott transition (metal-insulator transition, see Chapter 8) takes place. In these conditions, outer electron shells of neighboring impurity atoms overlap very slightly. Each donor exists therefore as an isolated hydrogen-like atom whose outer electron is at a distance about a_B^* from the nucleus and whose bonding energy is about E_B^* (see Chapter 8). Assume that the temperature of the semiconductor is so low that the thermal energy of vibrating atoms is not sufficient to ionize a donor. What could be the mechanism of electric conduction in this situation?

Imagine that some donors lost their outer electron. Usually this results from impurity compensation (see Chapter 10). If the semiconductor contains both donor and acceptor impurities, each acceptor accommodates one electron from a donor. If the number of acceptors is smaller than that of donors, a fraction of donors will have retained an outer electron, while the other fraction will have lost their electrons and become positively charged.

The mechanism of hopping electric conduction consists in the "hops" of an electron from one donor to another which before this hop had no outer electron.

Now we are going to consider the case of an outer donor electron whose potential energy depends only slightly on the spatial position of the donor, that is, on the specificities of the configuration of charged impurities surrounding this donor. This case corresponds to a low concentration of impurities. In conditions typical for hopping conduction, the spread in the energies of outer electrons belonging to different donors is roughly 0.1 of the bonding energy E_B^*.

In this situation, the attraction of an electron to the donor to which it belongs at the initial moment is the main obstacle to a "jump" from one donor to another. From the standpoint of classical mechanics, an electron can be transferred from the outer orbit of one donor to the outer orbit of another if some work is done against attractive forces because it is necessary to remove the electron to one half of the distance between the donors. After this point the electron is attracted to the second donor. If donors are spread very thinly, this work is almost equal to the work required to remove the electron belonging to an isolated donor to infinity, that is, equal to the bonding energy E_B^*.

What then is the advantage that hopping electric conduction holds over the electric conduction by free electrons? The point is that *hopping electric conduction is a quantum phenomenon*. Quantum mechanics allows an electron to pass from one donor to another without rising to a free state and without borrowing energy from the thermal motion of atoms. This is the *tunnelling transition*. Tunnelling transitions obey the law of energy conservation. The law imposes the constraint that the electron energy be equal in the initial and final states. Consequently, if the energies of the electrons at the first and second donors differ owing to the potentials of the surrounding impurities, a deficit of energy has certainly to be borrowed, and a surplus of energy has to be dissipated. But this energy is ten times less than E_B^*. As a result, hopping electric conduction defeats the electric conduction by free electrons at very low temperatures.

Resistor Network

It must not be overlooked, however, that a tunnelling transition is a very low-probability event in conditions we discuss. As was said in Chapter 8, the probability of finding an outer electron of a donor at a distance r from the donor's nucleus decreases with r as $\exp(-2r/a_B^*)$. If two donors *1* and *2* are at a distance r_{12} from each other, the probability of finding the outer electron of donor *1* near the nucleus of donor *2* equals $\exp(-2r_{12}/a_B^*)$. This exponential enters the probability of tunnelling transition. As follows from Chapter 8, at a donor concentration much less than the critical concentration N_c of the metal-insulator transition, the mean distance between donors is much greater than a_B^*, so that the quantity $\exp(-2r_{12}/a_B^*) = 1/\exp(2r_{12}/a_B^*)$ is, as a rule, very small.

Nevertheless, tunnelling transitions do take place time and again between neighboring donors. If an electric field is applied to a semiconductor, these transitions will be more frequent in the direction against the field (along the applied force) than along the field. The result is an electric current proportional to the electric field strength. This is the essence of the phenomenon we call hopping conduction.

The model used to calculate the resistivity of a semiconductor is the so-called "resistor network". This model is formulated not in terms of atoms and tunnelling transitions but of conventional electric circuit resistors. Imagine that a resistor is connected between each pair of donors. The donors themselves can be pictured as tiny metallic balls to which the wires from numerous resistors are soldered. The second end of each of these resistors is soldered to another ball. The result should resemble an irregular three-dimensional resistor network. Obviously, an intention to assemble an actual network model would require that the scale of the system be substantially magnified. Indeed, the mean distance between donors is of the order of 10^{-5} cm.

The resistors connecting two donors must be found by calculating the tunnel current that flows between these donors at a given electric field. This means solving a quantum-mechanical problem that will not be given here. Note only that, in view of the above description, the electric current produced by tunnelling transitions between donors in a given field is the smaller, the greater the separation between these donors is.

Accordingly, the resistance $\mathscr{R}$ connecting donors at a distance r from each other can be written in the form

$$\mathscr{R}(r) = \mathscr{R}_0 \exp(2r/a_B^*) \qquad (1)$$

where $\mathscr{R}_0$ is identical for all resistances (it can be of the order of 1 ohm). The problem is to find the resistivity of a system composed of a tremendously large number of donors (10^{16}-10^{19}) with ran dom patial distribution.

Properties of Resistor Network

The main feature of the model we discuss is that the resistances described by formula (1) are spread over a fantastically large range. The mean distance r_{don} between donors is related to the donor concentration N_{don} by $(4/3)\,\pi r_{\mathrm{don}}^3 N_{\mathrm{don}} = 1$, which means that on the average one donor is nested in a sphere of radius r_{don}. Typically, hopping electric conduction is observed in conditions in which r_{don} exceeds a_{B}^* by a factor of 6 to 12.

Assume, for instance, that $r_{\mathrm{don}} = 10a_{\mathrm{B}}^*$. Then the resistance connected between donors at a distance $1.5r_{\mathrm{don}}$ is greater than that between donors at a distance r_{don} by a factor $\exp(r_{\mathrm{don}}/a_{\mathrm{B}}^*) = e^{10} = 2.2 \cdot 10^4$.

Donors at a distance r_{don} occur almost as frequently as those at a distance $1.5r_{\mathrm{don}}$. Consequently, a negligible change in separation between donors results in a tremendous variation of resistance connected between them.

In principle, the resistor network model stipulates that each pair of donors is connected through a resistor. However, the resistances between remote donors are so large that they can undoubtedly be neglected. As a rule, the same two donors are connected through a chain of resistors connecting nearest-neighbor donors. Although the length of this chain is greater than the shortest distance between remote donors, the chain resistance is much less than one resistance connecting these two donors. Such are the properties of the exponential function: if $x_1 \gg 1$ and $x_2 \gg 1$, then $e^{x_1+x_2} = e^{x_1} \cdot e^{x_2} \gg e^{x_1} + e^{x_2}$;

Consequently, it is amply sufficient to retain in the resistor network only those resistors that connect each donor with four or five of its nearest neighbors.

The Sphere Problem Revisited

Our next step is to calculate the resistances. We suggest the following line of reasoning. Let us disconnect all resistors supposed to join the balls that stand for donors and start connecting them in a predetermined sequence.

First we solder in the resistances that connect the donors separated by a distance less than a certain length r'. This means that we include the lowest resistances smaller than $\mathscr{R}' = \mathscr{R}_0 \exp(2r'/a_B^*)$. If r' is much smaller than the mean separation r_{don} between donors, then at this stage we connect only the rare donors separated by anomalously short distance. As a rule, these resistances are not interconnected and cannot sustain the flow of current through the system.

Let us gradually increase r', each time adding new groups of resistances. Beginning with a certain value $r' = r_c$, the resistors form an infinite cluster. At $r' > r_c$ the system is electrically conductive and has a finite resistivity.

The problem of finding r_c is nothing less than the sphere problem (see Chapter 7). Indeed, let us construct a sphere of radius r' around each donor. Then resistances will be connected only between those donors one of which lies inside the sphere constructed around the other. According to the results of Chapter 7, such donors join into an infinite cluster when the condition

$$B_c = (4/3)\,\pi N_{\rm don} r_c^3 = 2.7 \pm 0.1$$

is satisfied, whence

$$r_c = (0.865 \pm 0.015)\, N_{\rm don}^{-1/3} \approx 1.39 r_{\rm don} \tag{2}$$

Calculation of Resistivity

As was mentioned in Chapter 12, the backbone of an infinite cluster can be pictured as a three-dimensional wire network (see Fig. 44) with knots spaced by distances of the order of the correlation radius. In the present problem each wire must be regarded as composed of numerous balls connected with resistors, and expression (1) of Chapter 12 must be rewritten in the form

$$R(r') = r_{\rm don} \left(\frac{r_{\rm don}}{|r' - r_c|} \right)^{\nu} \tag{3}$$

Expression (3) turns to infinity at percolation threshold, with the critical exponent ν which, by virtue of universality, is the same as in other three-dimensional problems of percolation theory ($\nu = 0.8$-0.9). When $r' - r_c$ is of the order of $r_{\rm don}$, we are far from the threshold, and the correlation radius becomes of the order of $r_{\rm don}$.

When $r' = r_c$, the separation between network sites in an infinite cluster grows to infinity. At precisely this point, the density of the infinite cluster is still zero. However, if $r' > r_c$, the cluster forms channels going across the whole system and resulting in a finite resistivity.

Let us continue with the procedure of adding resistances. If we switch on resistances between donors spaced by distances from r_c to $r_c + g a_B^*$, where g is a number smaller than unity (e.g. $g = 0.2$), the maximum one among the connected resistances will remain practically unaltered because $\exp[(2r_c/a_B^*) + 2g] \approx \exp(2r_c/a_B^*)$. On the other hand, an infinite cluster represents a network with site-to-site separation of about

$$R_c = r_{\text{don}} \left(\frac{r_{\text{don}}}{g a_B^*}\right)^{\nu} \tag{4}$$

In the theory of hopping conduction this network is referred to as *critical*. It is through this network that the electric current flows.

Indeed, resistivity will not be appreciably changed by any further increase of r', that is, by adding the remaining resistances. As we find from formula (1), resistances for $r - r_c \gg a_B^*$ are very large in comparison with resistances for $r = r_c$. In fact, practically no current flows through these resistances since it "prefers" the critical network whose maximum resistances are

$$\mathscr{R}_{\max} = \mathscr{R}_0 \exp(-2r_c/a_B^*)$$

The last step to accomplish is to calculate the resistivity of a critical network. By analogy to what we did in Chapter 12, this network will be represented by a wire skeleton modelling a simple cubic lattice (see Fig. 12) with period R_c.

Each wire connecting two neighboring sites of the lattice consists of a large number of resistances given by formula (1) and connected in series. As individual resistances differ from one another enormously, the resistance of one wire must be set equal to the maximum one of the resistances it comprises, that is, to $\mathscr{R}_{\max}$. The cubic unit cell of a simple cubic lattice (see Fig. 27) consists of four wires connected in parallel, each wire belonging at the same time to four other cubic unit cells, so that there is one wire with resistance $\mathscr{R}_{\max}$ per each unit cell.

In order to find resistivity, it is necessary to multiply the total resistance $\mathscr{R}_{\max}$ of a cubic unit cell by the area of its face (R_c^2) and divide it by the edge length (R_c). This gives

$$\rho = \mathscr{R}_{\max} R_c = \mathscr{R}_0 \cdot R_c \exp(2r_c/a_B^*) \tag{5}$$

We thus deal with a cubic unit cell as if it were filled by a homogeneous material with resistivity ρ. This is the meaning imparted to the concept of resistivity of a system that is in fact strongly inhomogeneous.

Discussion of the Result

It must be borne in mind that different elements of formula (5) were obtained with different degrees of accuracy. The factor $\mathscr{R}_0 R_c$, called the *preexponential factor*, was found to within a numerical coefficient. Indeed, we do not really know the quantity g in formula (4) for R_c. Besides, the replacement of the network of an infinite cluster by a periodic lattice may well lead to an error in the numerical coefficient. It must be mentioned that *no theory has been constructed* yet that could reliably determine the numerical coefficient in the preexponential factor.

As for the quantity in the exponent of the exponential expression (5), it is known to a good accuracy. In the true spirit of the derivation of formula (5), r_c in the exponent could be replaced by $r_c + a_B^* g$.

Then the sum $2r_c/a_B^* + 2g$ would appear instead of $2r_c/a_B^*$. The second term of the sum actually characterizes the uncertainty in our knowledge of the exponent. As has been mentioned, $2r_c/a_B^* \gg 1$ at low donor concentrations, so that the relative error is small. In fact, it reflects the uncertainty in the numerical coefficient of the preexponential factor.

The dependence of resistivity ρ on a donor concentration is determined solely by the exponent of the exponential. By virtue of (2) and (5),

$$\ln \rho = \ln \mathscr{R}_0 \cdot R_c + \frac{2 \cdot 0.86}{N_{\text{don}}^{1/3} a_B^*} \tag{6}$$

The first term on the right-hand side of (6) depends on N_{don} much weaker than the second term, and thus can be considered constant. A comparison between formula (6) and experimental data, carried out for numerous semiconductors, demonstrated that this formula gives a very good fit to the $\ln \rho$ vs N_{don} dependence; this agreement was an important achievement of the theory described above.

The method discussed in this chapter was successfully applied to determining the temperature dependence of hopping electric conduction, and to its dependence on the external magnetic field and on a number of other parameters. Furthermore, this method holds for any inhomogeneous system whose resistance varies in a wide range as a function of coordinates.

Chapter 14

Final Remarks

In this last chapter we shall briefly discuss some applications of percolation theory that have not been covered in the preceding chapters but nevertheless appear to us as quite interesting, and we shall also make a summary, in an attempt to point at the features common for all the problems discussed in the book.

Some Applications

Flow of liquid through a maze. This problem is closely linked to that of gas penetration into carbon gas masks that served as the starting point for percolation theory. Imagine a porous body through which a liquid is forced under pressure. The liquid does not wet the material of the body so that capillary forces resist the penetration of the liquid into pores.

A maze of capillaries permeates the body, but the capillary diameters are vastly different. They are wide in some places, and very narrow in others.

First imagine that the body has a single cylindrical capillary. At a certain pressure produced by a piston *2* a liquid is on the left of the body *1*, and at atmospheric pressure air is on its right (Fig. 48). The body is fixed and cannot move. Because of surface tension, a nonwetting liquid penetrates the capillary only if its pressure exceeds atmospheric pressure by the quantity $2\sigma/R$, where R is the capillary radius, and σ is the surface tension. At lower pressures, the liquid forms a convex meniscus but cannot pass through the capillary.

Now imagine that the capillaries within the body *1* are of different radii. Some of them (the widest) let the liquid flow through at a given pressure, but others (narrower ones) do not. As pressure increases, the number of permeable capillaries grows. At small pressures, when only the widest capillaries are accessible to the liquid, it cannot penetrate the body deeper than its subsurface layer. However, at a certain critical pressure, the permeable capillaries form a system that permeates the whole body. Beginning with this pressure, the liquid can be squeezed through the body.

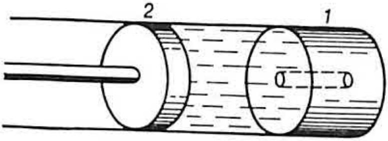


Fig. 48.

Percolation theory calculates this pressure as well as some other characteristics of the process, which constitutes a very important practical application of the theory.

Formation of polymer gels. A polymer consists of a very large number of elementary units (monomers). Monomers in a solution may bind to one another and form a complex three-dimensional network that permeates the whole system.

The result is a *gel*, that is, a solid-like medium resembling jelly.

A model is known that adequately describes the formation of a gel. It is essentially a problem of percolation theory in which white sites are the molecules of a monomer, and black sites are the molecules of a solvent. The bonds between white sites form with a temperature-dependent probability. A gel appears when an infinite cluster forms out of connected white sites.

This problem of percolation theory is said to be mixed because both sites and bonds are random elements here. Let the probability for a site to be white be x_1 (it equals the concentration of monomer molecules), and the probability for a given bond to be intact be x_2. We want to find the range of values of x_1 and x_2 in which an infinite cluster of interconnected white sites is formed.

By definition, x_1 and x_2 vary within the interval from zero to unity.

If $x_2 = 1$, that is, if no bonds are broken, an infinite cluster exists for all x_1 in the range $x_s \leqslant x_1 \leqslant 1$, where x_s is percolation threshold of the site problem. If $x_1 = 1$, that is, if all sites are white, the condition under which an infinite cluster is formed is $x_b \leqslant x_2 \leqslant 1$, where x_b is percolation threshold of the bond problem.

The square in Fig. 49 is the domain of the variables x_1 and x_2. The solid curve is the graph of the function $x_{\min}(x_2)$ that describes the boundary of the domain in which an infinite cluster exists. The function $x_{\min}(x_2)$ gives the minimum value of x_1 for each value of x_2 within the interval $x_b < x_2 < 1$, at which the infinite cluster exists.

It is readily understood that $x_{\min}(1) = x_s$, and $x_{\min}(x_b) = 1$. The domain of the infinite cluster in Fig. 49 is the hatched area.

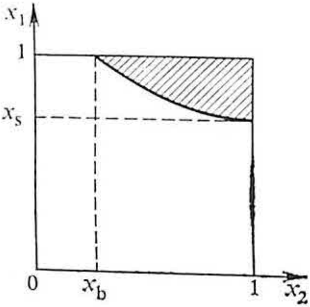


Fig. 49.

If the probability x_2 is known for each value of temperature, the function $x_{\min}(x_2)$ allows us to determine the range of temperatures and monomer concentrations in which a gel is formed.

What Is Percolation Theory, After All?

So far we have carefully avoided defining percolation theory. This would be far from simple. But having come to the last page of the book, we shall try to recapitulate the common element of all the problems outlined in the book, and formulate the subject of percolation theory.

Percolation theory deals with the connectivity of a very large (macroscopic) number of elements under the condition that the bonding of each element to its neighbors is random but prescribed in a well-defined manner (e.g. by a random-number generator with known properties).

The common element of percolation theory problems is that all of them have identical geometry of bonded elements in the vicinity of percolation threshold. In order to recognize this, it is necessary to disregard the small-scale structure depending on the type of bonding and the properties of elements, and monitor only the connectivity of large blocks. The universal large-scale geometry dictates the universal properties of physical quantities that depend on the structure of large clusters. This is the feature that unites percolation theory problems that look so dissimilar at first glance.

Answers and Solutions

Chapter 1

1. The general rule for calculating the mean value states that each possible value of a random variable must be multiplied by the probability of this value, and the products be summed up. The probability for each of the faces of the cube to be on top equals 1/6. Therefore, $\overline{a} = (1/6)\cdot 1 + (1/6)\cdot 2 + (1/6)\cdot 3 + (1/6)\cdot 4 + (1/6)\cdot 5 + (1/6)\cdot 6 = 21/6$

2. The results of individual trials change, but the average value $x_{\mathrm{c}}(\mathscr{N})$ calculated over many trials remains the same because on the average the rightward motion realizes with the same probability as the downward motion. Correspondingly, $x_{\mathrm{c}} = \lim\limits_{N\to\infty} x_{\mathrm{c}}(\mathscr{N})$ will not change either.

3. Let us denote the threshold values recorded in the ith trial by x_i' and x_i, where x_i' corresponds to the new definition, and x_i to the old definition (rightward percolation). If a decrease in x first interrupted the rightward percolation and then the downward percolation, then $x_i' = x_i$. But if the sequence was reverse, then $x_i' > x_i$. The average value of percolation threshold of the new definition is

$$x_{\mathrm{c}}'(\mathscr{N}) = \frac{x_1' + x_2' + \ldots + x_Q'}{Q}$$

and that of the old definition is

$$x_{\mathrm{c}}(\mathscr{N}) = \frac{x_1 + x_2 + \ldots + x_Q}{Q}$$

The total number of trials, Q, is considered very large in both formulas. However, situations with $x_i' > x_i$ will inevitably realize if the number of trials is large. For this reason, $x_c'(\mathscr{N}) > x_c(\mathscr{N})$. Nevertheless,

$$\lim_{\mathscr{N}\to\infty} x_c'(\mathscr{N}) = \lim_{\mathscr{N}\to\infty} x_c(\mathscr{N}) = x_c$$

The thing is that in an infinite system percolation threshold is not a random variable but a certain quantity, not varying from one trial to another. At the same time, differences between percolation thresholds for different directions represent a random phenomenon. From the standpoint of percolation, all directions are on the average equivalent. Hence, x_c is independent of direction.

4. The solution is quite similar to that of Problem 3.

Let us denote the threshold values obtained in the trial by x_i'' and x_i, with x_i'' corresponding to the new definition of threshold, and x_i to the old definition. It can be readily proved that $x_i'' \leqslant x_i$. Arguments similar to those of the preceding problem give

$$x_c''(\mathscr{N}) < x_c(\mathscr{N}), \quad \text{but} \quad \lim_{\mathscr{N}\to\infty} x_c''(\mathscr{N})$$

$$= \lim_{\mathscr{N}\to\infty} x_c(\mathscr{N}) = x_c$$

5. Formula (8) yields $\delta = 0.01$. This signifies that "typical" deviations from the mean value are ± 0.01. Hence, the last numeral in the estimate 0.59 reported by Watson and Leath was very likely in error. The probability of error in the first numeral after the decimal point is much smaller. Since only one trial was run, the authors of the paper could only evaluate the error with which they determined percolation threshold on the utilized sequence of blocked sites (it proved to be ± 0.005). However, they could not say anything about the possible change in the result if the experiment were rerun with a different random sequence. Later studies in which many trials were run at one value of $\mathscr{N}$, and studies with large values of $\mathscr{N}$, led to formula (8). These studies also demonstrated that even the second decimal place is correct in the number 0.59; to a large extent, this must be regarded as pure "luck".

Chapter 2

1. The fraction of blocked sites is $1 - x = \mathcal{N}'/\mathcal{N}$, and the fraction of nonblocked sites is $x = (\mathcal{N} - \mathcal{N}')/\mathcal{N}$. If Q sites are chosen at random, Qx of them will be nonblocked, and $Q(1 - x)$ will be blocked (the greater is Q, the better the accuracy is with which this relation holds). Consequently, the probability for a randomly selected site to be blocked is $Q(1 - x)/Q = 1 - x$, while the probability for it to be nonblocked equals $Qx/Q = x$. The sum of the probabilities equals unity because a site can be either blocked or nonblocked: $1 - x + x = 1$.

2. The probability of any sequence of three fixed numbers equals $1/6 \cdot 1/6 \cdot 1/6 = 1/216$. The number of different sequences satisfying the formulated conditions for the numbers 1, 2, 3 equals 6 (123, 213, 321, 231, 132, 312), and for the numbers 1, 2, 2 it is 3 (122, 212, 221). The probability for one of the possible sequences to realize equals the sum of these probabilities. Therefore, in the first case the sought probability is $6 \cdot 1/216 = 1/36$, and in the second it is $3 \cdot 1/216 = 1/72$.

3. $(0.8)^3 \cdot (0.9)^4 = 0.336$.

4. The solution is left to the reader.

5. The distribution function of a random variable a is constant within the interval $(-1, 1)$ and equals zero

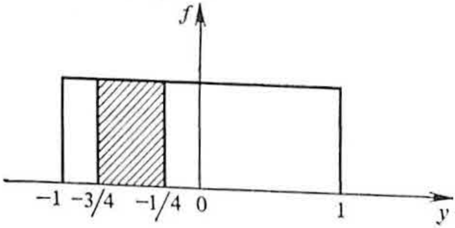


Fig. 50.

everywhere else (Fig. 50). The total area of the rectangle bounded by the curve $f(y)$ (in the present case by a horizontal straight line), the abscissa axis, and the perpendiculars erected at points -1 and 1 must equal unity; consequently, $f(y) = 1/2$ for $-1 < y < 1$. By virtue of the general rule, the sought probability equals the area of the rectangle bounded by the line $f(y)$, the abscissa axis, and the perpendiculars erected at points $y = -3/4$ and $y = -1/4$ (the rectangle is shaded in Fig. 50). The probability equals

$$[-1/4 - (-3/4)] \cdot 1/2 = 1/4$$

6. The variable y assumes all values from $-\infty$ to $+\infty$. Consequently, we have to set $A = -\infty$, $B = +\infty$ in formulas (3) and (4). According to formula (3),

$$\bar{a} = \int_{-\infty}^{\infty} y f_{\mathcal{N}}(y)\, dy$$

Substituting here formula (6), we find

$$\bar{a} = \frac{1}{\sqrt{2\pi}}\, \delta_{\mathcal{N}}^{-1} \cdot \int_{-\infty}^{\infty} y \exp\left(-\frac{y^2}{2\delta_{\mathcal{N}}^2}\right) dy$$

The integrand contains an odd function. Substituting $y = -t$ and comparing the result with the original formula, we note that $\bar{a} = -\bar{a}$, hence, $\bar{a} = 0$.

According to formula (4), the variance is

$$\delta^2 = \int_{-\infty}^{\infty} y^2 f_{\mathcal{N}}(y)\, dy$$

$$= \frac{1}{\sqrt{2\pi}}\, \delta_{\mathcal{N}}^{-1} \int_{-\infty}^{\infty} y^2 \exp\left(-\frac{y^2}{2\delta_{\mathcal{N}}^2}\right) dy$$

Substituting $y = \sqrt{2}\delta_{\mathcal{N}} \cdot t$, we obtain

$$\delta^2 = \frac{2}{\sqrt{\pi}}\,\delta^2_{\mathcal{N}} \int_{-\infty}^{\infty} t^2 \exp(-t^2)\,dt$$

The integral over t equals $\sqrt{\pi}/2$, so that $\delta^2 = \delta^2_{\mathcal{N}}$.

Chapter 3

1. By definition, $P(1) = 1$. If x is nearly unity, the sites may not belong to an infinite cluster for two reasons:

(i) They may contain nonmagnetic atoms. The fraction of such sites is $1 - x$.

(ii) Magnetic atoms may be isolated from an infinite cluster as, for example, atom B in Fig. 9. But if x is close to unity, so that nonmagnetic atoms are few, this factor is not so important because such isolation requires several nonmagnetic atoms to gather around one atom (four in the case of the plane lattice shown in Fig. 9). The probability of such an event is low if the number of nonmagnetic atoms is small. The second factor can thus be safely neglected and we can assume that the fraction of atoms belonging to the infinite cluster simply equals the fraction of magnetic atoms. We thus have $P(x) = x$, provided $1 - x \ll 1$.

2. Each atom of a simple cubic lattice has six nearest neighbors located along the cube's edges (see Fig. 12). The probability W_0 for all the nearest neighbors of an atom to be nonmagnetic equals the product of six probabilities: $W_0 = (1 - x)^6$. The probability W of at least one of the atoms being a magnetic atom equals

$$W(x) = 1 - W_0 = 1 - (1 - x)^6$$

By virtue of formula (2) of Chapter 3,

$$P_2(x) = xW(x) = x\,[1 - (1 - x)^6]$$

so that if $x \ll 1$,

$$P_2(x) \approx 6x^2$$

Obviously,

$$P_2(x) = x[1-(1-x)^z]$$

for any lattice in which each atom has z nearest neighbors, so that for $x \ll 1$

$$P_2(x) = zx^2$$

3. Figure 51 shows 12 atoms in the neighborhood of atom *0*. All of them can take part in forming a three-atom cluster. Such a cluster can be composed of, for instance, atoms *1*, *0*, *2* if all these three atoms happen to be magnetic. The probability of this event equals the product of three probabilities: $x \cdot x \cdot x = x^3$. The probability for the cluster to be formed of atoms *0*, *4*, *12*, or of any other trio of atoms, also equals x^3.

First we have to answer the question about the number of such groups of three atoms. We begin with counting how many of these groups include atoms *0* and *1*. There are six of them: *015*, *018*, *019*, *103*, *102*, *104*.

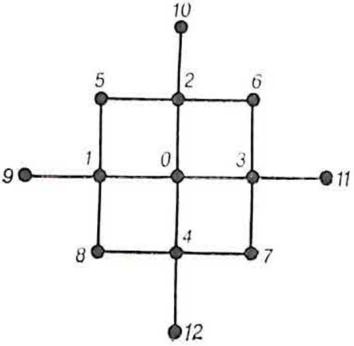


Fig. 51.

Now we consider three-atom groups including atoms *0* and *3*, but not atom *1*. There are five of them: *036*, *03 11*, *037*, *203*, *304*.

Likewise, there are four three-atom groups with atoms

0 and *2*, but without atoms *1* and *3*: *025*, *026*, *02 10*, *204*, and three-atom groups with atoms *0* and *4*, but without atoms *1*, *2*, *3*. These are *047*, *048*, *04 12*.

There are thus $6+5+4+3=18$ three-atom groups, each having the probability x^3. We want to find the probability for at least one of them to realize. If $x \ll 1$, the events consisting in the formation of any one of the three-atom groups can be treated as incompatible. Indeed, the probability for the groups *102* and *015* to form simultaneously equals the probability of *four atoms*, *0125*, being magnetic, that is, the probability of formation of a *four-atom* cluster. The probability of this event equals $x^4 = x^3 \cdot x \ll x^3$. Therefore, the incompatibility of these events is confirmed for $x \ll 1$. The probability for at least one of the three-atom groups to form is then equal to the sum of their probabilities, and

$$P_3 = 18x^3$$

4. The probability for a randomly selected atom to belong to a cluster of not less than two atoms can be represented as the sum of the probabilities of incompatible events:

$$P_2(x) = P_3(x) + \overline{P}_2(x) \qquad (1)$$

where $\overline{P}_2(x)$ is the probability for an atom to belong to a cluster of two atoms. Therefore,

$$P_3(x) = P_2(x) - \overline{P}_2(x) \qquad (2)$$

The function $P_2(x)$ is given by formula (3) of Chapter 3, and hence, we only need to calculate $\overline{P}_2(x)$.

Atom *0* (see Fig. 51) can form a two-atom cluster with atoms *1*, *2*, *3* or 4. The probability of the cluster being formed of atoms *0* and *1* equals the probability for both these atoms to be magnetic times the probability for atoms *2*, *3*, *4*, *8*, *9*, and *5* to be nonmagnetic, that is, it equals $x^2(1-x)^6$. The events consisting in the two-atom cluster being formed of atoms *0* and *2*, or *0* and *3*, or *0* and *4* are essentially the same. All these events are incompatible, and thus the probability $P_2(x)$ equals the sum of four probabilities

$$P_2(x) = 4x^2(1-x)^6 \tag{3}$$

Substituting formula (3) into formula (2), we obtain

$$P_3(x) = x[1-(1-x)^4] - 4x^2(1-x)^6 \tag{4}$$

which is the solution of the problem.

Making use of the binomial theorem, we can easily show that expression (4) contains no terms with power below 3. If $x \ll 1$, then $P_3(x) \approx 18x^3$, in accord with the result of the preceding exercise.

Chapter 4

1. 0.0085, 0.0072, 0.0051, 0.0026, 0.0006, 0.0000, 0.0000...; 0.0067, 0.0044, 0.0019, 0.0003, 0.0000, 0.0000...; 0.0032, 0.001, 0.0001, 0.0000, 0.0000....

2. The number b is in fact an n-digit number. Consequently, $b < 10^n$. In order to generate the next number b', we have to find b^2, divide it by 10^n, and take its integral part. Therefore, $b' < b^2/10^n$. But $b^2/10^n = b \cdot (b/10^n)$. Furthermore, $(b^2/10^n) < b$ because $(b/10^n) < 1$. We thus conclude that $b' < b$, which was to be proved.

3. 5, 15, 5, 15, 5, 15....

4. 5, 16, 9, 8, 5, 16, 9, 8, 5....

5. 5, 17, 13, 1, 5, 17, 13, 1, 5....

6. Condition (c) is nowhere satisfied. In Exercise 3, $c = 0$, in Exercise 5, condition (a) is not satisfied either, and so forth.

7. Let $X_0 = 0$. This gives us 0, 3, 1, 4, 2, 0, 3, 1, 4, 2, ..., and X_1 coincides with one of the numbers of this sequence whatever X_0 is.

8. The distribution function of random numbers obtained from the generator is shown in Fig. 4. These numbers form the array V. The fraction of nonblocked sites in the array V equals the fraction of random numbers in the array V, which satisfy the inequality $V < t$. There-

fore, the mean fraction of nonblocked sites equals the probability for a random number to be less than t. By definition of distribution function, this probability equals the area bounded by the curve $f(y)$, the abscissa axis, and two perpendiculars erected at points 0 and t. In the present case this is the area of a rectangle, equal to t. Hence, the mean fraction of nonblocked sites x equals t.

Chapter 5

1. When x is close to unity, almost all sites belong to an infinite cluster. Only those sites stay out of this cluster whose bonds connecting them with the surrounding system are all broken. The probability that one specific bond is broken equals $1 - x$. If the lattice is square, it is necessary to interrupt four bonds originating from a site in order to isolate this site from the system (Fig. 52*a*). The probability of this event equals the product of the appropriate probabilities, that is, it equals $(1 - x)^4$. In order to isolate two sites from the system, six bonds must be interrupted (Fig. 52*b*). The probability of this event is $(1 - x)^6$. If $(1 - x) \ll 1$, this probability is much smaller than that of one isolated site. Therefore, we can assume in the limiting case which is of interest now that all isolated sites are single, and the probability for a randomly selected site to be isolated is $(1 - x)^4$. The probability for a randomly selected site to be nonisolated equals $1 - (1 - x)^4$, that is, in a square lattice

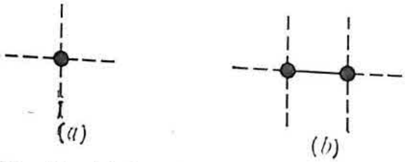


Fig. 52. (*a*) One isolated site; (*b*) two isolated sites. The intact bond is shown by the solid line, and the broken bonds by the dashed lines.

it is

$$P^{\mathrm{b}}(x) = 1 - (1 - x)^4$$

Similar arguments show that in a triangular lattice

$$P^{\mathrm{b}}(x) = 1 - (1 - x)^6$$

and in a honeycomb lattice

$$P^{\mathrm{b}}(x) = 1 - (1 - x)^3$$

These results hold if $(1 - x) \ll 1$.

2. The problem is solved by analogy to the preceding one. If $(1 - x) \ll 1$, almost all sites belong to an infinite cluster. A site chosen at random is isolated from the infinite cluster if all its nearest neighbors are nonmagnetic atoms (for definiteness, we use the terminology of the ferromagnetic problem). As in the preceding problem, the probability for a site to have for one of its neighbors a magnetic atom isolated from the infinite cluster is low. Therefore, it is sufficient to calculate the probability for all neighbors of the given site to be nonmagnetic. The probability that a site is occupied by a nonmagnetic atom is $1 - x$. The number of nearest neighbors equals the number of bonds that originate from a given site. Consequently, the results are the same as in Problem 1:

in a square lattice: $P^{\mathrm{s}}(x) = 1 - (1 - x)^4$
in a triangular lattice: $P^{\mathrm{s}}(x) = 1 - (1 - x)^6$
in a honeycomb lattice: $P^{\mathrm{s}}(x) = 1 - (1 - x)^3$

Hence, $P^{\mathrm{s}}(x) = P^{\mathrm{b}}(x)$ for $(1 - x) \ll 1$, in agreement with formula (2).

3. Consider a honeycomb lattice with a fraction x of white bonds, and a triangular lattice with a fraction y of white bonds. (We remind the reader that the term "white bond" is synonymous to "unbroken bond", and "black bond" is synonymous to "broken bond". The phrase "a site is connected with another site" must be interpreted, unless expressly stated otherwise, as equivalent to "a site connected through unbroken, i.e. white, bonds".) Let us superpose these lattices as shown in Fig. 53. Such sites as A, B, C are common for the two lattices, while

sites such as D belong only to the honeycomb lattice. The idea of the arguments that follow lies in that the percolation problem on a honeycomb lattice is reducible to the corresponding problem on a triangular lattice.

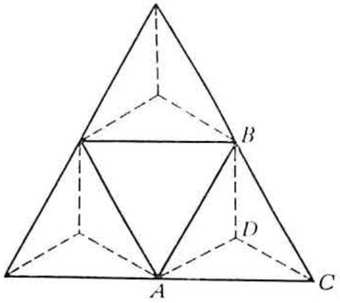


Fig. 53. Star-delta transformation. The dashed lines show the honeycomb lattice, and the solid lines show the triangular lattice.

The probabilities for sites A, B, C to be connected with one another must be expressed in terms of the fraction x of white bonds in the honeycomb lattice. We need to make use of the geometry and statistical properties of the bonds that originate from D-type sites. After this we can discuss only the triangular lattice shown in Fig. 53 by the solid lines and forget that each triangle has the dashed lines and D-type sites.

This operation, known as the star-delta transformation, is frequently used in designing electric circuits.

In fact, we shall need the following quantities:

(1) W_{--} is the probability for site A not to be connected with either site B or site C. This probability equals the sum of the probabilities of two incompatible events. The first event consists in bond AD being black, with arbitrary colors of bonds BD and DC. Its probability equals $1 - x$. The second event consists in bond AD being white, and both bonds BD and DC being black. The probability of this event equals the product of three

probabilities, $x(1-x)(1-x)$. As a result,

$$W_{--}(x) = 1 - x + x(1-x)^2 \tag{1}$$

(2) W_{+-} is the probability for site A to be connected with B but not with C. It equals the probability of bonds AD and DB being white, and bond DC being black, and is found as the product of the probabilities of all three events:

$$W_{+-} = x^2(1-x) \tag{2}$$

(3) W_{-+} is the probability for site A to be connected with C but not with B. It is readily found that $W_{-+} = W_{+-}$.

(4) W_{++} is the probability for site A to be connected with both B and C. It equals the probability for all the three bonds AD, DB, DC to be white and is calculated as the product of their probabilities:

$$W_{++}(x) = x^3 \tag{3}$$

With these four probabilities known, we can leave aside the honeycomb lattice and solve our percolation problem on the triangular lattice. If this problem is solved, we shall have the critical value x_b (H) for the honeycomb lattice.

Of course, this transformation does not in the least facilitate the solution. However, the same probabilities W can be expressed in terms of the fraction y (it represents the fraction of white bonds on the triangular lattice). At percolation threshold these probabilities assume quite definite values, as yet unknown; however, equating the probabilities expressed as functions of x to the probabilities as functions of y, we can derive a relation between the thresholds of the honeycomb (x_b (H)) and triangular (x_b (T)) lattices.

Our next problem is thus to derive all four probabilities as functions of y.

(1) W_{--}. For site A to be unconnected with both B and C, the color of bond BC is immaterial, but bonds AB and AC must be black. The probability of this event equals the product of probabilities:

$$W_{--}(y) = (1-y)^2 \tag{4}$$

(2) W_{+-}. Site A is connected with B but unconnected with C only if bond AB is white and *both* bonds BC and AC are black:

$$W_{+-} = y\,(1-y)^2 \tag{5}$$

(If e.g. bond BC were white, site A would connect with C via path ABC.)

(3) $W_{-+} = W_{+-}$, as in the preceding case.

(4) W_{++}. The probability for site A to be connected with both B and C equals the sum of the probabilities of four incompatible events. The first event consists in all the three bonds AB, BC, AC being white. Its probability is y^3. The other three events consist in only one among the three bonds being black. For instance, if the black bond is AB, then site A is connected with C via a white bond AC, and with B via path ACB. The probability of each of the three events equals $y^2\,(1-y)$. Finally, we obtain

$$W_{++}(y) = y^3 + 3y^2\,(1-y) \tag{6}$$

At percolation threshold all $W(x)$ must be equal to the respective $W(y)$. This yields a system of equations

$$W_{--}(x) = W_{--}(y);\ \ 1-x+x\,(1-x)^2 = (1-y)^2 \tag{7}$$

$$W_{+-}(x) = W_{+-}(y);\ \ x^2\,(1-x) = y\,(1-y)^2 \tag{8}$$

$$W_{++}(x) = W_{++}(y);\ \ x^3 = y^3 + 3y^2\,(1-y) \tag{9}$$

These equations must be satisfied by $x = x_b\,(\mathrm{H})$ and $y = x_b\,(\mathrm{T})$. Furthermore, percolation thresholds satisfy relation (23) of Chapter 5, which yields that $x_b\,(\mathrm{H}) = 1 - x_b\,(\mathrm{T})$.

The substitution of

$$y = x_b\,(\mathrm{T}),\ x = 1 - x_b\,(\mathrm{T})$$

into Eqs. (7), (8), and (9) transforms (8) into an identity,

$$(1 - x_b\,(\mathrm{T}))^2\,x_b\,(\mathrm{T}) = x_b\,(\mathrm{T})\,\,(1 - x_b\,(\mathrm{T}))^2$$

and (7) and (9) into the identical cubic equation:

$$x_b^3\,(\mathrm{T}) - 3x_b(\mathrm{T}) + 1 = 0$$

This equation has a single root within the interval $0 \leqslant x_b\,(\mathrm{T}) \leqslant 1$:

$x_b\,(T) = 2\sin(\pi/18) \approx 0.347\ 296$

Correspondingly,

$x_b\,(H) = 1 - x_b\,(T) = 1 - 2\sin(\pi/18) \approx 0.652\ 704$

4. The first step is to find the area per site for each lattice shown in Fig. 16, under the condition that the spacing between nearest neighbors is a.

Square lattice. Each unit cell includes four sites, but each of these sites belongs to four adjacent unit cells. Hence, there is one site per each square unit cell; in other words, the area per site is that of one square, that is, $S\,(S) = a^2$.

Triangular lattice. Each unit cell includes three sites, but each of these sites belongs to six adjacent unit cells. Hence, we find half a site per each triangular unit cell, and the area of two triangles per each site. The area of an equilateral triangle with side a is $\sqrt{3}a^2/4$. Therefore, the area per site is $S\,(T) = \sqrt{3}a^2/2$.

Honeycomb lattice. Each hexagon includes six sites, but each of these sites belongs to three adjacent unit cells. Hence, there is one hexagon per two sites. The area of a hexagon equals six times the area of an equilateral triangle with side a, that is, equals $3\sqrt{3}a^2/2$. Hence, the area per each site is $S\,(H) = 3\sqrt{3}a^2/4$.

The period a is determined in each lattice via a predetermined function $a\,(x)$, with percolation threshold x_b of the appropriate lattice used for x. This gives

$$S\,(T) = \left(\sqrt{3}/2\right)[a\,(0.35)]^2$$

$$S\,(S) = [a\,(0.5)]^2$$

$$S\,(H) = \left(3\sqrt{3}/4\right)[a\,(0.65)]^2$$

Obviously, the function $a\,(x)$ decreases with increasing x. (If the trees infect each other easier, the separation within a pair correspondingly decreases.) It is readily noticed, however, that this proposition is not sufficient in itself to derive even one of the above-written inequalities for areas. Indeed, in the honeycomb lattice the period a is the shortest, but the numerical coefficient in the expression S (H) for area is the greatest, while in the triangular lattice the situation is reverse. The lattice with minimum area thus cannot be found unless we know the function a (x) in greater detail.

Chapter 6

1. The filling factors for plane lattices are easily calculated from the results of Exercise 4 of Chapter 5. We have illustrated this in the text for the honeycomb lattice. We shall, therefore, restrict the analysis to 3D lattices, or rather to two of them, leaving the remaining checks to the reader.

Simple cubic lattice. Each cubic unit cell (see Fig. 12) contains eight sites, but each of the sites is shared by eight adjacent unit cells. Hence, there is one site per one unit cell, and the volume per each site equals the volume of the cubic unit cell, i.e. a^3. The spheres drawn around the sites have a radius $a/2$ and a volume $4\pi a^3/24$. The fraction of volume occupied by the spheres equals the ratio of the volume of one sphere to the volume per site. Hence, in a simple cubic lattice

$$f\ (\mathrm{SC}) = 4\pi/24 \approx 0.52$$

Body-centered cubic lattice. Each cubic unit cell (see Fig. 28*b*) contains nine sites, eight of which are at the vertices of the cube and one is at the center point. The site at the center belongs to one cubic unit cell exclusively, while each of the vertex sites belongs to eight adjacent unit cells. Hence, there are two sites per one unit cell, and the volume per each site is half the volume of the cubic unit cell, i.e. $a^3/2$. The nearest neighbor of each site is at a distance of half the body diagonal of the unit cell, i.e. $\sqrt{3}a/2$. The radius of a sphere drawn around each site is $\sqrt{3}a/4$ and the volume of the sphere is $(4\pi\cdot 3^{3/2}/3\cdot 64)\ a^3 = (\pi\sqrt{3}/16)\ a^3$. The filling factor equals the ratio of the volume of the sphere to the volume per site, so that

$$f\ (\mathrm{BCC}) = \pi\sqrt{3}/8 \approx 0.68$$

Chapter 7

1. The first coordination group of site *0* consists of 12 type-*1* sites (Fig. 54), the second consists of six type-*2* sites, and the third of 24 type-*3* sites (not all of them are shown in the figure); $12 + 6 + 24 = 42$.

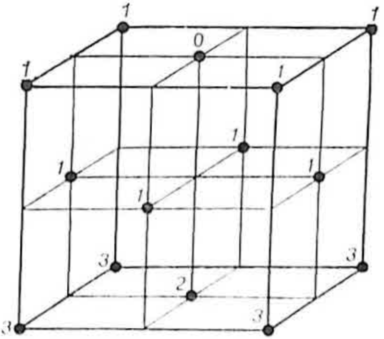


Fig. 54. Neighbors of site *0* in the fcc lattice.

Chapter 8

1. From formulas (3) and (5) we find $a_B^* = 636\cdot10^{-8}$ cm $= 636$ Å, $N_c = 7.8\cdot10^{13}$ cm^{-3}. This is a very low critical concentration. The transition to metallic conduction takes place when there is one impurity atom per 10^8 atoms of the semiconductor matrix! A very sophisticated technique of purification is required to manufacture semiconductors with such a low impurity concentration.

Chapter 11

1. Let us introduce a function $Q(x)$, defining it as the probability for an information that has appeared at a randomly chosen site to reach only a finite number of other sites. As before, $P(x) = 1 - Q$. The number of independent channels through which the information emerges from each site is q. Let us determine the probability for one of these channels to be interrupted at some step. The interruption can result from one of two incompatible events, a or b: a consists in the first bond of the channel being broken, and b consists in the first bond being intact, but the site to which this bond leads being capable of passing the information to only a finite number of other sites. The probability of event a is $1 - x$, and that of event b is $xQ(x)$. The probabilities of incompatible events are added. Consequently, the probability for one channel to be interrupted equals $1 - x + xQ(x)$. All the channels being independent, the probability for all of them to be interrupted is $[1 - x + xQ(x)]^q$.

This gives an equation for $Q(x)$:

$$Q(x) = [1 - x + xQ(x)]^q$$

This equation reduces to Eq. (4) by the substitution $Q' = 1 - x + xQ$. Finally, we obtain

$$P = \frac{(x - 1/q)\,2q^2}{q-1}$$

Note that percolation thresholds in the site and bond problems on the Bethe lattice are identical ($x_c = 1/q$). In fact, this could be predicted at the start. Indeed, let us assume that the site to which the broken bond leads is blocked. This allows us to assume that all bonds are intact, so that our bond problem has been transformed to the site problem. Hence, these problems necessarily have identical percolation thresholds.

Chapter 12

1. We need to calculate the resistance of a cube with unit-long edge. The number of wires connected in parallel is, as before, $1/R^2$, but the length of each wire does not any more equal one unit of length. It is longer by a factor

equal to the ratio $(\mathcal{L}/R) = (x - x_c)^{-(\zeta-\nu)}$. Correspondingly, the resistance of one wire is not ρ_0 but $\rho_0 (x - x_c)^{-(\zeta-\nu)}$. The necessary result is therefore obtained by replacing the quantity ρ_0 in formula (6) for σ_3 by $\rho_0 (x - x_c)^{-(\zeta-\nu)}$. The substitution gives

$$\sigma = \sigma_3 (x - x_c)^{\zeta+\nu}$$

where $\sigma_3 = \rho_0^{-1} l^{-2}$. Hence, $t = \zeta + \nu$.

To the Reader

Mir Publishers would be grateful for your comments on the content, translation and design of this book. We would also be pleased to receive any other suggestions you may wish to make.

Our address is:
Mir Publishers
2 Pervy Rizhsky Pereulok
I-110, GSP, Moscow, 129820
USSR

Printed in the Union of Soviet Socialist Republics

Other Titles in This Series

G. I. Kopylov, D.Sc. (Phys.-Math.)

Elementary Kinematics of Elementary Particles

This book tells a fascinating story of one of the basic goals of physics today: the discovery of the primary building blocks of matter. This field of science is called particle, or high-energy, physics, and is one of the frontiers of present-day physical research.

How is it possible to detect particles a hundred thousand times smaller than the atom, which itself is as many times smaller than an apple as the apple is smaller than the earth? How can we follow the motion of particles that have a velocity almost that of light? How can we measure the lifetime of these particles when it is of the order of 0.0000000000000000000000001 second? What kind of a clock can we use? How can we investigate the properties of these astonishing and elusive bits of matter?

All these and many other questions are comprehensively answered in this book written for the layman by the late Dr. Gertsen Kopylov, who was a prominent scientist, well known in the world of particle physics. This book was written to be understood even by readers having only a secondary school education and it requires a knowledge of only elementary algebra and geometry. Nevertheless, in the author's treatment, the material is in no way oversimplified or distorted.

Ya. A. Smorodinsky

Temperature

This book starts with a historical background on the notion of temperature and the development of the temperature scale. Then Ya. A. Smorodinsky covers the fundamentals of thermodynamics and statistical physics, only using concepts that will be familiar to high-school students. Having built a solid foundation, he exposes the reader to a number of phenomena that are essentially quantum-mechanical, but for which the concept of temperature "works", and works very well. These include the spins in crystal lattices, inverse population of energy levels, microwave background radiation, black holes, and cooling antiproton beams.

Although it has been written for high-school students, the book contains a minimum amount of mathematics. Nevertheless, Ya. A. Smorodinsky compensates for this severe restriction by the lucid manner in which he discusses very complicated effects.

L. V. Tarasov
A. N. Tarasova

Discussions on Refraction of Light

The rainbow and the Galilean telescope, the spectre of the Brocken and illuminated fountains—what can all these have in common? The answer is that in all of them the refraction of light has an important role to play. That is why you will read about them in this book written by the well-known popular science authors Lev and Aldina Tarasov. Professor Tarasov is a prominent figure in the field of quantum electronics and optics; Aldina Tarasova is a teacher of physics and studies methods for teaching the basics of science.
The book treats refraction of light from two points of view. Firstly, it describes the many, and sometimes quite surprising, forms that the refraction of light can assume. Secondly, the book traces the long and difficult path mankind covered before it came to understand some of the mysteries of Nature. Recent inventions such as the laser and the transmission of images over optic fibres are also dealt with. The book was first published in Russian in 1982, and has come out in English. We are certain it will be well received by those who are interested in the history of science and the present development of theoretical optics and its related technology as well as by all lovers of Nature.

Th. Wolkenstein

Electrons and Crystals

The increasingly important field of solid-state physics concerns the behaviour of electrons in various crystals. Problems of solid-state physics, which include specific differences between metals and dielectrics and the remarkable properties of semiconductors, are particularly topical in today's 'electronic' society.

Electrons and Crystals by Dr. Theodore Wolkenstein covers the fundamentals of solid-state physics in an engaging way. Written in an easy, readable style, the book is intended as a supplement to textbooks in secondary-school physics courses, and the approach to certain topics in the volume is, therefore, unique. The material is presented in terms of models and requires no special additional knowledge. Suitable for the general reader with a good command of elementary physics and mathematics, this book can also serve as a useful study guide for high-school students.

S. C. B. R. T. W.
Library
Calcutta

513

SCIENCE FOR EVERYONE

This book is about percolation theory and its various applications, which occur mostly in physics and chemistry. The book is self-sufficient in that it contains chapters on elementary probability theory and Monte Carlo simulation. Most attention is paid to the relationship between the geometrical and physical properties of systems in the vicinity of their percolation thresholds. The theory is applied to examples of impurity semiconductors and doped ferromagnetics, which demonstrate its universality. Although written for students at high schools, the book is very good reading for college students and will satisfy the curiosity of a physicist for whom this will be a first encounter with percolation theory.